Author:
Humphreys, Jane

Title:
WIRDS, WERDS, WYRDZ: Visual wordlikeness, lexical phonology, and models of visual word recognition

WIRDS, WERDS, WYRDZ: Visual wordlikeness, lexical phonology, and models of visual word recognition

Jane Humphreys

A dissertation submitted to the University of Bristol in accordance with the requirements of the degree of Doctor of Philosophy in the Faculty of Social Sciences and Law.

Department of Experimental Psychology, January 2008.

58,359 words

Abstract

Nonwords such as *brate* and pseudohomophones such as *brane* are often used to explore the processes involved in decoding orthography, without the potential confound from semantics. Experiments using these useful items can provide evidence that sheds light on competing models of visual word recognition in general and the status of lexical phonology in particular. However, previous experiments have often used stimuli that look implausible as exemplars of English spelling, (e.g. *phret, woez*), and it is arguable that some of the current controversies in the area may be partly attributable to the use of such stimuli. To investigate this notion, new items were constructed from real words on the grounds that they would contain a high proportion of existing orthotactic patterns. Ratings were gathered for the visual wordlikeness of the previous stimuli and these new items; the latter generated higher ratings than the former. Analysis of the ratings suggested that readers are sensitive to multiple sources of orthographic and graphophonemic information. In a series of naming and lexical decision experiments using the new stimuli, results showed that participants responded to visual wordlikeness across all tasks; for example, reading wordlike pseudohomophones more quickly than unwordlike, and responding to them more slowly in visual lexical decision. A masked priming experiment using wordlike and unwordlike primes showed that lexical phonology was less likely to be activated for the unwordlike pseudohomophones than the wordlike. Overall, the results support a view of visual word recognition as a highly-interactive system, processing multiple grain-sizes of sublexical and lexical information in which phonology plays a functional, non-optional, role. While orthotactic violations constrain its normal workings, the system has mechanisms that can be used to process unwordlike items; but it is unlikely that these processes are the same in all respects as those used for wordlike stimuli.

Acknowledgements

I gratefully acknowledge the patient and helpful guidance of my supervisor, Dr Clive Frankish.

I would like to thank the Department of Experimental Psychology, University of Bristol, for financial assistance.

I am grateful to my colleagues at the University of the West of England for helping me find the time to carry out and write up the research.

I thank my husband, sons, sons' partners, family, friends, and colleagues for their unfailing encouragement, especially at those times when the path appeared lost and the incline impossibly steep.

Author's declaration

I declare that the work in this dissertation was carried out in accordance with the Regulations of the University of Bristol. The work is original, except where indicated by special reference in the text, and no part of the dissertation has been submitted for any other academic award. Any views expressed in the dissertation are those of the author.

SIGNED: .... DATE: 22/05/08

Contents

Tables and illustrative material

"I have a little book, much briefer than Aristotle or Ovid, in which is contained the whole of science, and with very little study one may form from it the most complete ideas. It is the alphabet, and no doubt anyone who can properly join and order this or that vowel and these or those consonants with one another can dig out of it the truest answers to every question, and draw from it instruction in all the arts and sciences."

(Discourses, Galileo)

Chapter one

The role of phonology in visual word recognition

"'Twas brillig and the slithy toves
did gyre and gimble in the wabe
all mimsy were the borogoves
and the mome raths outgrabe."
(Lewis Carroll, Alice through the Looking Glass)

Most adults decode print into sound and meaning with unthinking ease, and, as demonstrated in the lines above, not only can we assign probable pronunciations to unknown words, we can also put limits on to their possible meanings: *slithy* describes *toves*, who appear to be sentient beings doing something somewhere. But, unlike spoken language, this skill is only acquired by instruction and with effort, and for many children, mapping sound and meaning to written symbols presents problems that may persist into adulthood. The facts that the skill is not easily acquired, and that written languages have only rarely occurred in human history (see DeFrancis, 1989), indicate that reading is a true accomplishment of human cognition. Understanding the processes involved can therefore not only help illuminate our view of basic cognitive processes, such as those involved in representation and computation, but also shed light on our depiction of the precise nature of the specific mechanisms involved. Additionally, understanding these processes may help provide a basis for understanding the difficulties faced by some children when learning to read. The problem of reading is that of extracting meaning at various levels (word, phrase, sentence); the fundamental requirement of reading is word recognition, and its study is one of the oldest areas of research in experimental psychology (Cattell, 1886). If word recognition processes do not operate fluently, reading will be impaired.

There are three tasks that have to be achieved; decoding the print symbols themselves, mapping sounds on the symbols, and mapping meaning on to the symbols. The second of these tasks, moving from print to sound, is a question that has been addressed by a large body of research over at least the past 30 years, and is the focus of this thesis. If such activation occurs even when it is not apparently necessary, for example in silent reading, it can be interpreted in two ways. Either it can be seen as an epiphenomenon of the system, possibly as a remnant of the way in which readers originally learn to map sounds to meanings, or it can be seen to be a mandatory, automatic process that reflects the workings of an essential component of the reading system. In this view, it is not possible to decode the meaning of print without invoking phonology; the former position holds that visual-orthographic processing is the primary process involved in reading, and that phonological activation is secondary. These two positions were usefully reviewed by Frost (1998), who made the case for a strong phonological approach to reading, as opposed to the prevailing view that skilled readers predominantly use visual-orthographic processing, bypassing computed phonology.

The work reported in the following chapters consists of a series of naming and lexical decision experiments designed to explore the processes involved in moving from print to sound, without reference to semantic representations. The appropriate stimuli are therefore novel letter strings. In order to explore the mechanisms involved in phonological activation, comparisons are made between pseudohomophones (such as *brane*) and true nonwords (such as *brate*). Both types of stimulus are orthographically meaningless, but the pseudohomophones have phonological lexical status whereas the true nonwords do not. If the pseudohomophones activate base word phonology, then differences in response times in naming and lexical decision experiments are predicted. The work also addresses the issue of stimulus construction, arguing that some of the areas of disagreement in the literature may be attributable to the use of unwordlike items. Such items, it is argued, may generate patterns of processing that are different from those used for more wordlike items. The experimental stimuli were therefore devised to be more orthotactically representative of English monosyllables; it is argued that responses to these items are potentially more informative about the print-to-sound decoding system than items that are atypical of those encountered in everyday reading. A final masked priming experiment suggests that wordlike stimuli do indeed activate different mechanisms from more unwordlike stimuli: that is, *wirds* and *werds* generate different responses from *wyrdz*.

1.1 Models of reading

Theoretical models of single word reading can be divided into two broad groups. The single-route approach proposes that reading all types of words and nonwords is accomplished by a single pathway from orthography to phonology. The dual-route approach maintains that known words are read via a system that already contains lexical representations, while nonwords are read by a system that uses grapheme-phoneme correspondences. Both approaches incorporate a semantic route, as shown in Figures 1 and 2, but the key difference between the models is whether one or two routes or processes are needed to derive sound from print, and the single- and dual-route terminology arises from this question of how to conceptualise the pathway from orthography to phonology. The single-route approach tends to be associated with a parallel distributed processing or connectionist approach while the dual-route formulation tends to be linked with the concept of a mental lexicon. There are of course exceptions: Zorzi, Houghton and Butterworth (1998) presented a connectionist model with two routes to deal with the pronunciations of nonwords and exception words. However, the approach that has dominated most of the literature over the past 30 years is the one that postulates a lexical route for known words and a nonlexical route for nonwords and its most clearly defined competitor has been the single-route model; these are the two that will be outlined in the following sections.

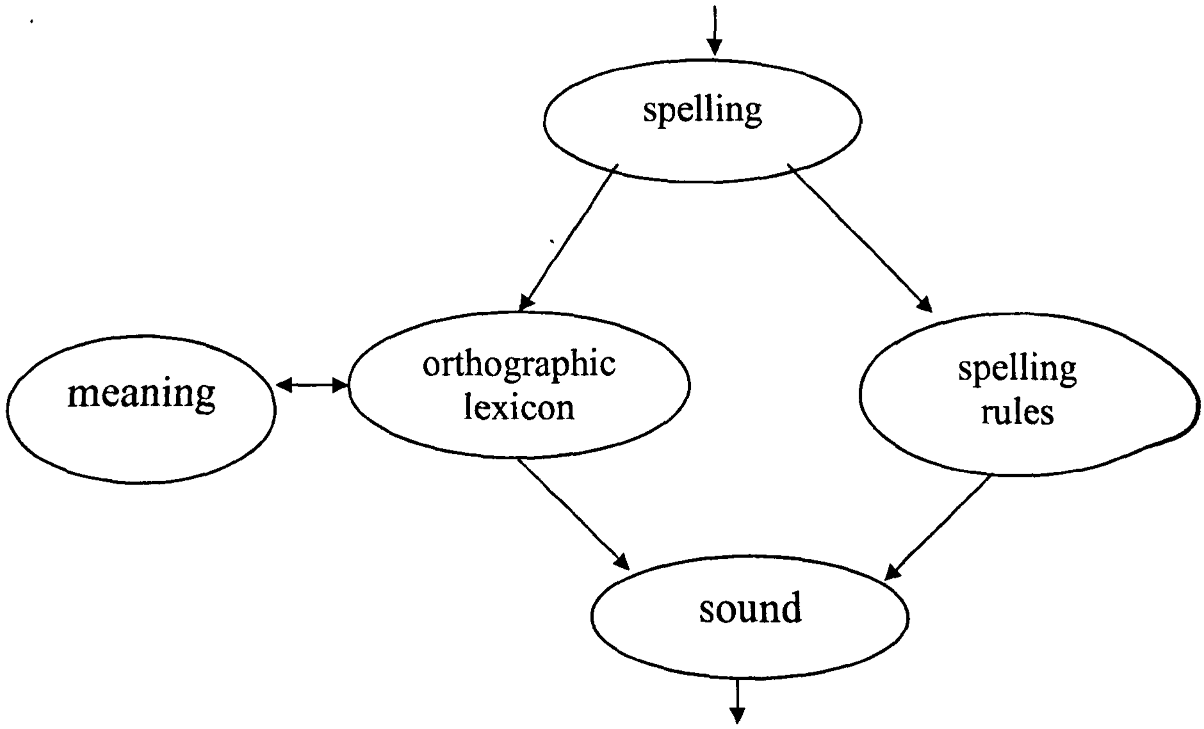


Figure 1. A dual-route model of spelling-to-sound.

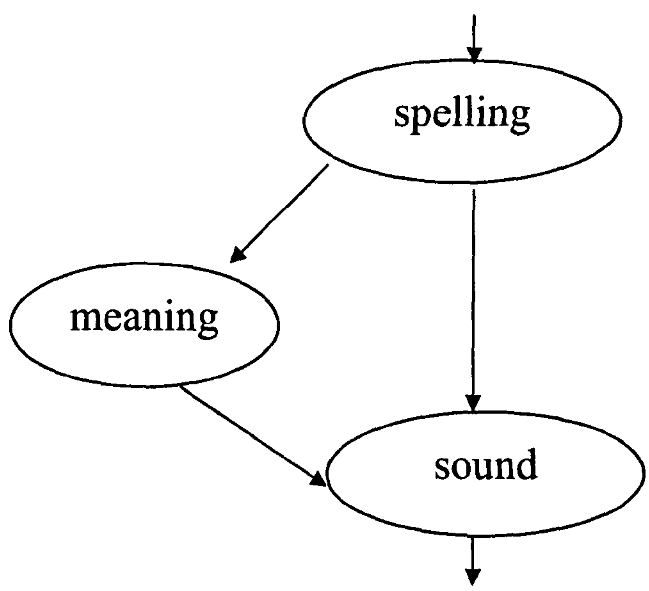


Figure 2. A single-route model of spelling-to-sound.

1.1.1 Dual-route models

In a dual-route approach, orthographic stimuli may be processed along two routes, which operate either together or separately. One route converts a stimulus into a phonological code by a rule-based process that maps letters, graphemes or larger units on to phonemes. This is known as the assembly, nonlexical, or grapheme-to-phoneme conversion route. The second route involves the direct access of stored orthographic and phonological representations; this is the lexical route. Information from both routes is passed on to an output buffer that resolves any conflict between the two codes, and then synthesizes these codes for articulation. In this type of model, new words must be processed using the assembly route, because there are no entries corresponding to these items in the lexical route. Known rules of orthography are used to assemble a pronunciation, with the result that an exception or irregular word such as *pint,* when met for the first time, might be pronounced as *hint,* following normal rules of pronunciation for *–int* i.e. with the /ɪ/ pronunciation for the vowel. Later, as *pint* becomes a known word, it will be added to the corpus of word representations accessed by the lexical route, and will be pronounced correctly.

Evidence for such a division of labour comes from studies of dyslexia, a reading impairment which may include a selective inability to read certain types of words. Cases of dyslexia appear to offer a classic "double dissociation" in that one form demonstrates impairment in the reading of exception words while nonword and regular word reading are spared, while the other form manifests impairment in nonword reading relative to reading of both regular and exception words. In the former case, known as surface dyslexia, there tends to be a high proportion of regularisation errors in reading irregular words (e.g. Marshall & Newcombe, 1973; McCarthy & Warrington, 1986; Patterson, Marshall, & Coltheart, 1985). In the latter case, known as phonological dyslexia, difficulty with nonword reading is shown (e.g. Beauvois & Derouesné, 1979; Berndt, Haendiges, Mitchum & Wayland, 1996; Funnell, 1983). The explanation for these two types of dyslexia in dual-route terms is that the lexical route is damaged in surface dyslexia while the nonlexical route is relatively spared, whereas in phonological dyslexia the nonlexical route is damaged but the lexical route is relatively spared.

However, evidence from cases of dyslexia is by no means clear-cut and, by itself, it does not offer entirely convincing evidence for the existence of two separate routes.

For example, it is sometimes difficult to label a case as being distinctly either phonological or surface; furthermore, cases are rarely identical. It is probable that the "purest" cases are the ones that enter the literature; Ellis, Lambon Ralph, Morris and Hunter in Funnell (2000) identified three sub-types of surface dyslexia in order to account for the different patterns of impairment and suggested that differences between cases may lead "one to seriously question whether a single explanation will ever account for all of the observed cases of surface dyslexia" (p. 97). Patterns of deficit may also vary depending on whether the dyslexia is developmental dyslexia, or acquired as a result of stroke or other brain insult. In the latter case, as Ellis et al. pointed out, pre-morbid individual differences may result in different patterns of impairment. Also, there could also be a variety of other impairments that might affect word reading. Opponents of the dual-route perspective argue that the dichotomy between phonological and surface dyslexia is unjustified by the evidence (e.g. Harm & Seidenberg, 1999).

One approach to putting a theoretical explanation to the test is to implement it in computational terms. A computational implementation forces theorists to be specific about the processes that might be occurring at all stages in the model. Another advantage is that if the model does not work, it may reveal ways in which the theory is incomplete or underspecified. Once the model does work, it can then be used to test the adequacy of the theory by simulating effects observed in experiments with humans. More importantly, models can predict phenomena that have not yet been observed, although such demonstrations are in practice rarely observed.
The most influential dual-route implementation in recent years has been the Dual-Route Cascaded (DRC) model (Coltheart, Rastle, Perry, Langdon, & Ziegler, 2001). This model has simulated a wide variety of empirical phenomena and has important implications for our understanding of single word reading. The semantic route is currently unimplemented; the lexical and nonlexical routes are fully-implemented. The early part of the lexical route is based on the principles embodied in an earlier computational model, the interactive activation and competition (IAC) model (Grainger & Jacobs, 1996; McClelland & Rumelhart, 1981; Rumelhart & McClelland, 1982). This has three representational levels: a visual feature level, a

letter level, and an orthographic word level. In addition, there is a phonological lexicon. Both orthographic and phonological lexicons are frequency-sensitive. Finally, there is a phoneme unit level which is activated by entries in the phonological lexicon. The lexical route passes information in cascaded fashion, as with the IAC, so that as soon as information is received at one level, it is passed on to the next level. The cascaded approach is preferred by these authors to the alternative threshold approach, because it can account for lexical effects in nonword processing (e.g. an inconsistent nonword like *heaf* takes longer to process than *hean* because of activation of the real-word *deaf*, cf. Glushko, 1979). The nonlexical route is based on the model described by Coltheart, Curtis, Atkins, & Haller, (1993), in which a rule-discovery algorithm learned a set of grapheme-to-phoneme rules from the exposure to the database of about 3,000 word spellings and their pronunciations compiled by Seidenberg and McClelland (1989). The 2001 model converts graphemes to phonemes by using inbuilt correspondence rules; this set of rules is thought of as a "set of hypotheses about what GPC rules people know" (Coltheart et al., 2001, p. 216). Initially, visual features and corresponding letter units are activated (a process performed in common with the lexical route), then the constituent letters of the stimulus are processed left-to-right and serially, and finally, when a pronunciation has been assembled, it is fed to the phoneme units. Information from the lexical and the nonlexical routes is pooled in order to produce an output; so nonword naming is, for example, facilitated by activation of orthographically similar words in the lexical route.

Coltheart et al. (2001) demonstrated that the DRC model was able to simulate 25 effects reported in studies of lexical decision and reading aloud. In lexical decision, for example, it simulated effects of word frequency and neighbourhood size (Andrews, 1989, 1992). In simulations of reading aloud, cycles to naming were shorter for high-frequency words than low-frequency words (Forster & Chambers, 1973); and words were faster to name than nonwords (McCann & Besner, 1987). When lesioned, the model demonstrated some "dyslexic" behaviours; for example, changing the GPC route's parameters to increase the number of cycles between accessing one letter and the next from 17 to 27 produced a pattern of results similar to that of phonological dyslexic patient LB (Derouesné & Beauvois, 1985). However, as pointed out by Plaut (1999), the architecture of the dual-route model provides for complete elimination of one route and complete sparing of the other, but such a pattern of impairment has never been clinically demonstrated.

Nevertheless, this model has to date simulated a wider range of single word reading phenomena than any other computational model and has been the focus of much subsequent research (e.g. Rastle & Brysbaert, 2006). This achievement should not be underestimated. However, the selection of experimental results to model may be biased by theoretical principles or computational possibilities, and the reporting of successful simulations may mask the fact that simulations using different stimuli may not have produced the same results. Moreover, just because a computer model produces the same sort of results as humans, this does not necessarily demonstrate that this is how humans do the task; indeed, this issue is acknowledged by Coltheart et al. (2001, p. 216), when discussing the rule-discovery algorithm they implemented for the GPC route. Therefore we still need to explore alternative ways of conceptualising the processes involved in single word reading, in both theoretical and computational terms. One important alternative approach is captured in single-route explanations.

1.1.2 Single-route models

The notion that there are distinct lexical and nonlexical routes has been strongly challenged by theorists who argue that all types of words and nonwords are read using a single route (e.g. Plaut, 1999; Plaut, McClelland, Seidenberg, & Patterson, 1996; Seidenberg & McClelland, 1989; Van Orden, Pennington & Stone, 1990). This type of account argues that, first, stored whole-word lexical representations are not a necessary element of visual word processing, second, inbuilt rules are not required, and third, all words and nonwords are processed via the same processes. An important characteristic of this approach is that "regular" and "exception" words are both read in a single-route system that has picked up on the statistical structure among written and spoken words and the contexts in which they occur. In this view, the strict dichotomy between "regular" items that obey rules and "exception" items that disobey the rules is redundant, a conclusion originally drawn in 1979 by Glushko; thus the alternative view is that words are, to a greater or lesser extent, consistent with the patterns of mappings between spellings, sound and meaning to which the reading system has been exposed.

A number of parallel distributed processing (PDP) computational models have been implemented to make this perspective concrete (e.g. Harm & Seidenberg, 1999; Plaut, McClelland, Seidenberg & Patterson, 1996; Plaut, 1999; Seidenberg & McClelland, 1989). These models, consisting of a large number of interconnected neuron-like units or nodes, aim to demonstrate that a single system can learn mappings between orthography and phonology by changes in the weights between nodes. Implemented models typically consist of input and output nodes, and one or more layers of hidden nodes. An example in the shape of Seidenberg and McClelland's influential framework is shown in Figure 3. Statistical correspondences that are inherent in exposure to print, such as word frequency and consistency of spelling-to-sound mappings, are captured by the pattern of weights connecting the units in the network. Knowledge of the relationships between orthographic input and phonological output is embedded in the system in terms of the weights that have been learned during previous exposure to orthographic stimuli. When exposed to new items, the system can apply this distributed, embedded knowledge to produce a phonemic output.

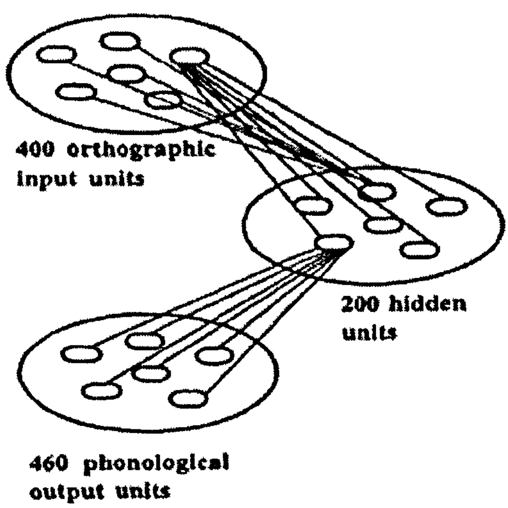


Figure 3. Single-route model of reading as implemented by Seidenberg and McClelland, 1989. (Reproduced from Coltheart, Curtis, Atkins & Haller, 1993).

These models have captured various phenomena associated with visual word recognition. One of the first networks designed to simulate reading showed that a single set of processes could deal with both regular and exception words, thus showing that it was unnecessary to invoke two discrete routes (Seidenberg & McClelland, 1989). This model demonstrated the frequency by consistency interaction in naming latency (Andrews, 1982; Seidenberg, Waters, Barnes, & Tanenhaus, 1984), but the model's performance was limited in that it was unable to generalise its knowledge to nonword naming. This problem was overcome in the Plaut et al. (1996) model, but successful nonword naming was dependent on employing localist input coding, in that the input scheme was based on larger graphemic units. However, the simulation reported by Plaut (1999) was based on letter-by-letter reading, and was also able to read both words and nonwords, as well as showing other effects such as word length and a frequency-by-consistency interaction demonstrated in human readers.

While the double dissociation of phonological and surface dyslexia is neatly explained by dual-route theory in terms of selective damage to one or the other route, it is less easy for single route theorists to explain. Zorzi et al. (1998) state that "dyslexic dissociations offer substantial challenges to any single-route account of reading" (p. 1134). Phonological dyslexia has been described as a "thorn in the flesh" (Funnell, 2000, p.10), and surface dyslexia as a "serious difficulty" (Coltheart, 2006, p. 101) for the single-route model. Zorzi et al.'s connectionist model is very like a traditional two-route model, with two separate routes, one for exception words and one for nonwords; lesioning the first route produced an inability to read exception words, as in classic cases of surface dyslexia (e.g. patients M.P., Bub, Cancelliere, & Kertesz, 1985, and K.T., McCarthy & Warrington, 1986). Clearly, support for a connectionist approach will be stronger if behaviours typical of both surface and phonological dyslexia can be shown to emerge from selective lesions to a triangle-type model, with a single orthography-to-phonology pathway. Harm and Seidenberg (2001) suggest that phonological dyslexia emerges as a result of impairment to phonological representations; they damaged the phonological attractors of their model and achieved results qualitatively similar to the behaviour of phonological dyslexic MJ in that her nonword reading was affected by graphemic complexity (Howard & Best, 1996). Surface dyslexia was modelled by Plaut et al., (1996). Impairment to the phonological pathway on its own did not produce surface dyslexia, but when the pathway was trained in the context of semantics, impairment to the semantic route did produce this behaviour. The suggestion is that the two pathways of the triangle model become co-dependent in training and that damage to the semantic pathway results in a decrease in competence of the phonological pathway, which in turn produces surface dyslexia. Plaut (1997) carried out additional simulations that suggest that premorbid differences in the divisions of labour between semantic and phonological pathways can account for individual differences in the severity of surface dyslexia. More recently, Welbourne and Lambon Ralph (2006) have simulated both surface and phonological dyslexia with simulations using Plaut et al.'s model; this is currently the only demonstration of the dissociation shown in a single PDP reading model.

It may be, however, that the category of surface dyslexia is unhelpful, at least in developmental dyslexia: Harm and Seidenberg (1999) have argued that it has emerged as part of the dual-route vernacular, but that because a discrete inability to read exception words is rarely seen, this category should be seen more as a general delay in the acquisition of reading skills. Evidence from statistical studies supports this view; for example, Stanovich, Siegel and Gottardo (1997) showed that while phonological dyslexia can be reliably identified as a deviant pattern of reading, surface dyslexia emerges as a form of developmental delay. In trying to simulate surface dyslexia with a connectionist network, Plaut and Shallice (1994) reported that damaging the network did not work as well as simply examining the undamaged network at an earlier point in its learning curve. Surface dyslexia, it seems, may just be a case of slow, but normal, learning. The delay might be attributable to a number of causes: such as a shortage of computational resources, lack of experience, or failures to learn efficiently from experience.

It appears, then, that PDP models are able to simulate some of the key findings in normal and impaired word recognition. However, some researchers have argued that the models are unconvincing; for example, Spieler and Balota (1997) correlated the mean naming latencies for 2,820 words with models' latencies for the same words and found that the models accounted for only about 3 – 10% of the variance associated with individual items. By contrast, the traditional measures of log frequency, orthographic length and neighbourhood size (measured as Coltheart's N; Coltheart, Davelaar, Jonasson, & Besner, 1977) collectively accounted for 21.7% of the variance; and including an encoding of the phonetic properties of the onset phoneme increased this to 43.1%. It may be that these results were attributable to the training regimes, especially in regards to how frequency is instantiated (see Plaut et al., 1996 for a discussion). A fundamental issue is raised here: how do we decide what is acceptable as an adequate replication of human data? While some researchers report a "qualitative" match of human and computational patterns of data (e.g. Harm & Seidenberg, 2001), others adopt a more quantitative approach by comparing statistical tests of human and computer results (e.g. Coltheart et al., 2001). It might be argued that the latter approach is more convincing; on the other hand, given the wide variability in human response times, it could equally well be argued that a single run-through of a connectionist model is simply behaving like a human – that is, idiosyncratically.

Arising from this is the more immediate point that matching models' behaviour to humans' is difficult if the empirical evidence is inconclusive in the first place. Although there are some incontrovertible phenomena relating to word and nonword reading, there are also areas where the evidence is not so clear-cut, particularly in nonword and pseudohomophone naming and lexical decision experiments. Dealing with nonwords is an important attribute of the reading system, since all words when encountered for the first time are nonwords, and effective systems are needed to decode them. According to the single-route approach, the ability to deal with nonwords emerges from previous experience with similar items; according to the dual-route model, nonword reading is primarily accomplished via the rules embedded in the serial, letter-by-letter grapheme-to-phoneme conversion route.

1.2 Words, wirps and wirds: varieties of letter strings

1.2.1 What do we know about *word* reading?

Any theoretical or computational model of reading has to deal with a range of phenomena that have been demonstrated with human readers. One such phenomenon is the so-called regularity effect. In English, letters and letter strings usually correspond to speech units (phonemes) in a predictable way, but this is sometimes irregular or arbitrary. For example, *pint* and *have* are not pronounced according to the majority of words that end in *-int* and *-ave*; but the pronunciation of *colonel* cannot be guessed from its spelling. Skilled English readers can usually read these words aloud accurately, and will also pronounce nonwords such as *kint*, *mave* and *solonel* according to standard phonological rules, rather than according to the exception word pronunciation. Such observations are neatly explained by dual-route models of reading. The lexical route deals with specific representations of all known words, including exception words, and therefore these words are pronounced correctly. The nonlexical route deals with explicit rules specifying the dominant pronunciation for letters and letter combinations, so the vowel sound in the nonword *mave* will be read as /eɪ/ rather than as /æ/ via this route, pronounced to rhyme with *save, gave* and *wave.* When a regular word is read, any input from the nonlexical route will not conflict with the lexical route's output, but when an irregular word is presented, the nonlexical route may provide an alternative conflicting pronunciation, which will slow the output. Most dual-route theorists would be in agreement with this broad-brush depiction of how we can account for exception word reading, but there is considerable disagreement about how the assembly and output processes actually work. Some models assume a 'horse race' between the two outputs (e.g., Paap & Noel, 1991) such that the response is determined by the first process to finish. Other models, including the DRC computational model, propose that the output of the two processes is pooled simultaneously until a phonological representation sufficient to drive articulation is generated (e.g., Monsell, Patterson, Graham, Hughes, & Milroy, 1992).

One problem that has already been mentioned is the characterisation of English spellings as either regular or irregular. It is difficult to see how the requirement that a word has to be one or the other predicates anything other than two discrete routes. But if spelling is delineated as a continuum where relationships are broadly consistent or inconsistent, this generates alternative ways of conceptualising the reading system, and a connectionist approach emerges naturally from this. Knowledge of spelling-to-sound mappings emerges from exposure to a body of print containing consistent and inconsistent items, and when a novel item is produced, it is quite likely that responses to consistent items will be more accurate than to inconsistent. Thus, the single-route model reported by Plaut et al. (Simulation 3, 1996) produced 'regular' pronunciations to 40 out of 43 (93%) of consistent nonwords, about equal to the 94% level of performance originally reported by Glushko (1979), and, like human readers, it performed less well for inconsistent nonwords.

Another area that has been extensively studied in word reading is the word frequency effect. Words that occur frequently in print are identified faster and/or more accurately than low frequency words (e.g. Andrews, 1989; Forster & Chambers, 1973; Scarborough, Cortese & Scarborough, 1977). In models where there are assumed to be whole-word representations, frequency effects are taken to reflect the rate at which activation of lexical representations has occurred (e.g. Coltheart et al., 1993, 2001; McClelland & Rumelhart, 1981) or rate of verification (Paap, McDonald, Schvaneveldt & Noel, 1987). In models in which separate word representations are not assumed, word frequency effects are manifest in the connection weights between the orthographic, phonological and semantic processing subsystems involved (e.g. Borowsky & Besner, 1993; Plaut et al., 1996; Zorzi et al., 1998). Additionally, reading aloud is faster for regular words than for irregular words when these are low in frequency; when they are high in frequency, the regularity effect is smaller or absent (Paap, Chen & Noel, 1987, Paap & Noel, 1991, Seidenberg, Waters, Barnes and Tanenhaus, 1984, Taraban & McClelland, 1987). Finally, reading has also been shown to be affected by N, (see Andrews, 1997, for a review), length (Weekes 1997), and the "whammy effect" (the finding that five-letter, three-phoneme words take longer to read than five-letter, five-phoneme words; Rastle & Coltheart, 1998). There is therefore a wide range of word reading phenomena that theoretical and computational models must be able to simulate. In addition, the ability to generate sensible pronunciations for nonwords must also be accounted for.

1.2.2 What do we know about *nonword* reading?

Nonword reading is of interest because each time a reader meets a new word, it is essentially a nonword, and therefore responses to this type of item can be studied to explore what happens when a reader meets a word for the first time. Experimental evidence paints a relatively clear picture of what happens when words are read, although interpretations of how it happens are less consensual. What readers do when they read nonwords is much less well delineated. To a certain extent, this may be because of an issue that is bypassed in most of the studies in this area, that of semantic representations. Usually when one meets a word for the first time in reading, it is in some kind of meaningful context. When nonwords are used as stimuli, the letter strings typically have no semantic representation, which allows an examination of decoding processes without the potential confound of semantic or morphological knowledge. However, this is essentially an unrealistic task and may account for some of the inconsistencies in the data. This issue needs to be acknowledged here, but left to one side for the time being in order to pursue a discussion along traditional lines of what might be happening based on the evidence from studies of 'pure' nonword reading.

Nonwords are useful stimuli because they enable us to explore how we move from print to sound in a relatively "pure" fashion, that is, without the confound of semantics and previous experience with the item. In all other respects it is clearly necessary that the stimuli should have properties that characterise real words, and these items are typically controlled for N, number of phonemes, bigram frequency and so on. Nonwords, not surprisingly, are named more slowly than words (McCann & Besner, 1987; Rastle & Coltheart, 1999). Naming is slower for nonwords containing digraphs and items constructed with non-existing orthographic rimes (Andrews, Woollams, & Bond, 2005).

An interesting sub-group of nonwords is that of pseudohomophones, which are essentially incorrect orthographic representations of legal phonological and semantic representations. For example, the pseudohomophone *brane* is a mis-spelling of the real word *brain*. It seems as though readers might respond differently to pseudohomophones than to true nonwords such as *brate*, and the interpretation of this evidence has important theoretical implications. Two phenomena that have been investigated are whether pseudohomophones are read aloud faster than true nonwords (McCann & Besner, 1987; Seidenberg, Petersen, MacDonald, & Plaut, 1996; Taft & Russell, 1992), and whether readers are sensitive to the frequency of the base words of the pseudohomophones, such that shorter latencies for more frequent items are elicited. Evidence from these stimuli is important because the dual-route model predicts speeded naming of pseudohomophones, plus a base word frequency effect, while the single-route model does not; so if we can establish the precise conditions under which the effects are obtained, this has important theoretical implications.

1.2.3 The pseudohomophone effect

When latency is the measure taken, pseudohomophones are usually read faster than control nonwords, (e.g. Coltheart, Laxon, Keating & Pool, 1986; Herdman, LeFevre & Greenham, 1994; Laxon, Coltheart, & Keating, 1988; Laxon, Smith, & Masterson, 1995; Marmurek & Kwantes, 1996, Expt 1; McCann & Besner, 1987; Seidenberg et al., 1996; Taft & Russell, 1992). One interpretation of the effect is that pseudohomophone processing is facilitated by stored base word representations in the mental lexicon (for example, the pronunciation of *brane* is facilitated by the stored phonological representation of *brain*). An alternative, single-route, explanation, would be that pseudohomophone effects can be attributed to the relative ease of articulating a phonologically familiar item versus an unfamiliar item (e.g. Herdman et al., 1994; Seidenberg et al., 1996).

The phenomenon, however, is not entirely robust. For example, Laxon, Masterson, Pool and Keating (1992) suggested that it is possible that the effect "comes and goes within the same lists depending on the measure taken (latency or pronunciation)" (p. 739). For the Taft and Russell (1992) findings, only the slow readers (n = 7) demonstrated a pseudohomophone advantage; Marmurek and Kwantes (1996) were unable to replicate this finding. It is possible that the McCann and Besner (1987) stimuli elicited a pseudohomophone effect because their stimuli began with more frequent onsets than the nonwords (see Seidenberg et al., 1996, Expt. 1). Laxon et al. (1992) suggested that uncontrolled orthographic factors were responsible for the mixed findings they reported. However, more recent attempts to address issues of stimulus composition have still generated conflicting accounts. For example, Seidenberg et al. (1996) reported strong pseudohomophone effects, in both delayed and immediate naming conditions, thus supporting their argument that the effect stems from post-lexical operations, but Grainger, Spinelli and Ferrand (2000) found the effect occurred only in immediate naming – the effect disappeared in delayed naming. The present state of play is that it is still unclear as to whether pseudohomophones are named faster than nonwords, or not; and this lack of clarity may at least in part be attributable to the orthotactic properties of the stimuli.

1.3 Phonological factors in pseudohomophone naming

Pseudohomophones are primarily of interest because of their phonological lexical status; if they access frequency-sensitive base word representations, a frequency effect would be expected, such that stimuli with more frequent base words would be named more quickly. Establishing the conditions under which this effect occurs is important not only because it can inform the single- versus dual-route debate, but also because it addresses the question of the nature of phonological activation. In his influential review, Frost (1998) described two positions: one of 'strong' phonology, which holds that phonological lexical representations are always and mandatorily activated, and a 'weak' position which contends that phonology may be activated but is not a default – the primary mode of access is visual-orthographic. The prevailing view in the literature is that a weak position best captures the empirical data (e.g. Coltheart et al., 2001; Harm & Seidenberg, 2004; Plaut et al., 1996; Zorzi et al., 1998) although the strong position has also generated support (e.g. Drieghe & Brysbaert, 2002; Lukatela & Turvey, 1994a; Van Orden, 1987).

The dual-route approach suggests that although processing of nonwords is primarily carried out by the nonlexical route, some support is given from activities co-occurring in the lexical route; this results in pseudohomophones being read faster than nonwords, and, if a lexical representation of the base word is activated, it would be logical to assume that responses will be sensitive to base word frequency. Thus, more frequent phonological representations should be named faster than less frequent ones. The pseudohomophone and base word frequency effect are harder for single-route accounts to explain, in that such models do not possess representations for whole words. However, single-route theorists explain the pseudohomophone and the base word frequency effects as occurring as part of the articulatory process, that is, after a pronunciation is decided upon. That is, *brane* is facilitated because articulatory mechanisms are familiar with *brain*, whereas *brate* does not have similar support.

Both types of models of single word reading would be better positioned to account for pseudohomophone naming and base word frequency effects if they co-occurred, but research has shown that they often do not co-occur and sometimes do not occur at all. The standard finding with skilled readers has been a pseudohomophone naming advantage (e.g., Herdman, LeFevre, & Greenham, 1996; McCann & Besner, 1987; Seidenberg et al., 1996) without a base word frequency effect. It has been a challenge for researchers to establish a convincing explanation of this finding.

1.3.1 Base word frequency

Early research suggested that there was no base word frequency effect on pseudohomophone naming latency, even though there was a large frequency effect for the base words themselves (McCann & Besner, 1987). McCann and Besner argued that because they obtained a frequency effect for the words but not for the comparable pseudohomophones, this offered a picture in which the orthographic lexicon was frequency-sensitive but the phonological lexicon was not. In the interests of parsimony, they argued that frequency effects were better characterised as emerging from the links that join the various components of lexical memory. Thus, word frequency effects are mediated by differences in the strength of connections between whole-word orthographic representations and output phonology, while pseudohomophones do not activate lexical orthography, and therefore do not have access to these whole-word links.

In any case, interpretations of the evidence are difficult because it is probable that McCann and Besner's stimuli, though constructed with a variety of controls according to procedures used at the time, were flawed. Indeed, the literature presents many examples where modifications to the stimuli and to the method of presentation lead to different results. For example, Seidenberg et al. (1996) concluded that orthographic properties such as N and bigram frequency confounded any real differences between pseudohomophones and control nonwords. Taft and Russell (1992) had earlier pointed out that a pseudohomophone's orthographic similarity to its base word might affect results. For example, *phocks* is orthographically so different from its base word *fox* it is unlikely to activate the lexical representation, and is more likely to be read via the nonlexical route than a pseudohomophone that is orthographically very similar. Based on this argument, they constructed pseudohomophones and nonwords in pairs, matching them on orthographic factors. Having found a base word frequency effect, they explained the results in terms of a traditional dual-route model. Since only slow readers demonstrated the effect, Taft and Russell argued that fast readers were using the rule-based nonlexical route and therefore did not show a frequency effect. Slow readers on the other hand, were using their knowledge of graphemes to access a likely orthographic match in the lexical route. Using their knowledge about which graphemes are normally pronounced in the same way (e.g. *-aim* and *-ame*), they were able to apply this knowledge to access the orthography of the base word. Thus the orthography of *gaim* and *daim* activates *game* and *dame,* but the latter takes longer to name because it is low frequency.

Only high frequency base words demonstrated the pseudohomophone effect in Taft and Russell's experiment, and it is likely that this was because they were more orthographically similar to words than the pseudohomophones in the low-frequency group. Pursuing this argument, Herdman et al. (1996) constructed a new set of stimuli controlling for summed positional bigram frequency (SPBF) and N across frequency categories, and found a pseudohomophone advantage but without a base word frequency effect, even though the base words themselves did exhibit a frequency effect. Borowsky, Owen, and Masson (2002) suggested that the only explanation for this is that the pseudohomophones were for some reason insensitive to frequency. They suggested that the reason might be that half of the pseudohomophones contained illegal orthographic rimes and this may have been enough to have obscured the effect – once again, the suggestion is that the findings were confounded by orthotactic properties of the stimuli.

Seidenberg et al. (1996) explicitly addressed problems of stimulus construction but still found a pseudohomophone effect without a base word frequency effect; they argue that this is attributable to the fact that while the pseudohomophone effect is a post-lexical effect, the base word frequency effect is not detectable because it is too fine-grained a measure. However, Borowsky and Masson (1999) examined Seidenberg et al.'s (1996) base words and found that they themselves did not elicit a reliable frequency effect, so perhaps it is not surprising that the pseudohomophones derived from these words did not show one either. Evidently, it is still not clear whether a pseudohomophone and base word frequency effect occurs in naming, and stimulus construction is a crucial consideration. Another related concern is that of list construction, because that too affects reader responses, but in a somewhat more reliable way.

1.3.2 List presentation

Mixed lists, using both pseudohomophones and nonwords, or with the addition of base words, have been the paradigm for much of the research cited (e.g. McCann & Besner, 1987; Taft & Russell, 1992; Herdman et al., 1996; Seidenberg et al.,1996). An alternative approach was adopted by Marmurek and Kwantes (1996) who argued that since frequency effects for words are dependent on list structure (e.g. Monsell et al., 1992), that factor probably also affects nonword naming. Thus, Marmurek and Kwantes found that when control nonwords were removed from McCann and Besner's list, a base word frequency effect did occur. They found that frequency effects were only apparent when pseudohomophones were presented in pure lists, and in lists containing words – when mixed with nonwords, the effect disappeared. Similarly, Grainger et al. (2000) found that a base word frequency effect for pseudohomophones presented on their own disappeared when pseudohomophones and controls were presented together. An explanation couched in dual-route terms is that the lexical route dominates in pure pseudohomophone lists, so that a baseword frequency effect occurs; but in mixed lists, the nonlexical route dominates and the baseword frequency effect is washed out. Since according to single-route theorists, there is no difference between nonwords and pseudohomophones in the orthography to phonology path, any difference between pure and mixed lists must be attributable to post-assembly operations, such as a lexical check in pure lists.

The position is further complicated by list presentation order. In a series of mixed and pure block experiments using previous stimuli and their own new items, Borowsky et al. (2002) found the usual findings of a pseudohomophone effect with no base word frequency effect in mixed lists, but different results in pure block presentations depending on whether nonword or pseudohomophone blocks were presented first. Generally, when pseudohomophones were presented first, there was a base word frequency effect plus a pseudohomophone *dis*advantage, but when nonwords were presented first, the effects disappeared (but not for the Seidenberg et al. stimuli, where the opposite results occurred). Borowsky et al.'s explanation for these results involved strategic processing and a general scaling effect, an account similar to McCann and Besner's in that it invoked a decoupling of the processes responsible for the pseudohomophone and the base word frequency effects.

In short, it is evident that different results emerge from pure and mixed lists, but what exactly these results are is still not clear. Additionally, it is possible that list composition effects are attributable to cognitive mechanisms that are separate from those involved in decoding print to sound (e.g. Lupker, Brown & Colombo, 1997). For example, it is likely that participants establish an appropriate response-time criterion when faced with naming a list of words. So rather than couching explanations in terms of route selection or de-emphasis, a more straightforward explanation might be that participants respond to the relative difficulty of the stimuli. Nonwords will elicit a different response criterion from pseudohomophones, so when these stimuli are in pure lists, different times will be given. But when they are in mixed lists, the criterion is set at an intermediate position, so that latencies tend to homogenize; and differences between the two types of stimuli are therefore no longer detectable.

1.4 Orthographic factors in pseudohomophone naming

Having looked at phonological factors, we need also to consider orthographic influences. First, this is a methodological issue; spelling of nonwords and pseudohomophones is a crucial consideration in interpreting results, and results of early studies were probably confounded by the fact that the pseudohomophones looked more like words than the nonwords, and that some pseudohomophones looked more like their base words than others (Martin, 1982; Taft, 2006). Second, it is an issue of theoretical interest; if we can equate pseudohomophones and nonwords in terms of their similarity to words, we can then go on to find out what, if any, additional orthographic variables are important as predictors of naming latencies. This in turn has implications for models of transforming print to sound; why, for example, might *skreme* elicit longer naming response times than *skream*? One explanation would be that the former takes longer because it looks less like its base word (Taft, 2006), while a different approach would argue that the final –e of *skreme* acts as a 'whammy' and delays sequential processing (Rastle & Coltheart, 1998).

1.4.1 Neighbourhood effects

The notion that nonwords can be orthographically more or less similar to real words is frequently captured by the concept of 'neighbourhood', or Coltheart's N, which is accorded an important status in symbolic accounts of word recognition (see Andrews, 1997, for a review). N is calculated by summing the number of real words generated if one letter is changed in a letter string. In the same way that the frequency effect is seen to be an indicator of phonological whole word contribution to pseudohomophone naming, so an effect of N is seen as an index of orthographic contribution to naming. The assumption is that letter strings will activate similar items, so that *bake,* for example, will activate *cake, lake, bike* and *bare.*

However, exactly how nonwords and pseudohomophone naming is affected by N is not entirely clear. McCann and Besner (1987) failed to find an influence of N on pseudohomophone naming, but for the nonword stimuli, items with few neighbours were slower and less accurate than those with many neighbours. Contradictory results were reported by Laxon et al. (1992) who observed N facilitation for both types of stimuli. Since word-naming latencies and nonword naming latencies are facilitated by increasing the neighbourhood size of stimuli (Andrews, 1989, 1992; Peereman & Content, 1995) it would be surprising if pseudohomophones were unaffected, so the McCann and Besner findings are again possibly attributable to some feature of pseudohomophone spelling. With stimuli controlled for orthography so that all pseudohomophones and all control nonwords were one letter different from their French basewords, (e.g. *sujet* - *suget* - *sunet*), Grainger, Spinelli and Ferrand (2000) obtained N effects for pseudohomophones when presented in pure lists, such that those with several neighbours were read faster than those with none. This effect interacted with baseword frequency, with N effects disappearing with high frequency stimuli, a pattern of results that mirrors performance to words. When pseudohomophones and controls were presented together in mixed lists, the baseword frequency effect for pseudohomophones disappeared, and N had a strong facilitatory effect on both types of nonword. These authors attributed the N effect to nonlexical processing; stimuli with several orthographic neighbours generally have more frequent spelling-to-sound correspondences than stimuli with just one neighbour. Coltheart et al. (2001), however, take a different view, stating that nonword N effects arise from orthographically similar words being activated in the orthographic lexicon.

1.4.2. Other orthographic effects

Some researchers acknowledge the view that N may be a spurious predictor variable (Andrews, 1997; Ziegler & Perry, 1998), and its apparent effect is actually attributable to the fact that it captures statistical regularities between sound and spelling. These regularities are likely to be particularly detectable in the rime. For example, Vanhoy and Van Orden (2001) found that N did not contribute to pseudohomophone naming latency, and, in lexical decision, accounted for only 1% of the variance once orthographic rime spelling was accounted for. Rime status also proved to be the salient predictor in nonword naming in Peereman and Content's (1997) exploration of neighbourhood effects in French. As suggested earlier, the relationship between the orthography of the base word and the pseudohomophone may also be important; with some research suggesting that the greater the orthographic similarity between pseudohomophone and base word, the faster the pseudohomophone is named (Marmurek & Kwantes, 1994; Spinelli, 1994). A useful metric in this context is 'orthographic similarity' (Van Orden, 1987; Vanhoy & Van

Orden, 2001), which is a measure of spelling similarity between base word and pseudohomophone, and captures the relationships between spelling and sound that are not accounted for in rime statistics.

How orthographic factors are conceptualised is closely linked to the nature of the model being explored. The idea of activation and inhibition of neighbouring words is implicated in models with lexical representations, such as the DRC model; those assumed to learn from exposure to a print-to-sound vocabulary are linked with the idea that some kind of statistical mappings will occur that represent the system's knowledge of orthotactic probabilities. N and other orthographic measures are likely to correlate, and we need to be clearer about what exactly the important orthographic features are in nonwords and pseudohomophones.

1.4.3 Stimulus construction

Researchers in this area have always been sensitive to the need to implement controls when constructing nonword stimuli, but notions of what factors need controlling have changed as the research enterprise has progressed. In some of the earliest studies, McCann and Besner (1987) and McCann, Besner and Davelaar (1988) controlled for mean bigram frequencies and N, as well as the possible confounds of ease of articulation and rimes. However, as has been pointed out by subsequent researchers, their nonwords were still somewhat heterogenous. For example, Laxon et al. (1992) noted that they varied from four to six letters, which, apart from anything else, affects N because N tends to decrease as word length increases. Also, the alterations to the base words were very varied; for example, changing *fox* to *phocks* is a different type of change from *first* to *furst.* In their follow-up study, Taft and Russell (1992) designed nonwords based on high and low frequency words in pairs, matching for orthography, so, for example, changing high frequency *game* to *gaim*, low frequency *dame* to *daim*, and generating nonword *raim*.

Laxon et al. (1992) devised high and low N pseudohomophones with only four letters, and took care to avoid unusual letter sequences. The alterations were mainly vowel changes because altering consonants produces orthographically very different nonwords, such as *phocks* and *kight.* All the nonwords had consistent endings, and controls were created from pseudohomophones by altering the first letter of the pseudohomophone. Onset frequencies, bigram frequencies, and N values were controlled for in Seidenberg et al.'s (1996) stimuli. Onsets and rimes of nonwords and pseudohomophones were swapped to produce matching quadruples such as *hoap*, *hoak*, *joap* and *joak*. Grainger el al. (1996) chose five-letter words selected according to frequency and N, and changed one letter to make a pseudohomophone or nonword (e.g. *genou/jenou/menou*). With such controls in place, the assumption is that the letter strings are valid experimental stimuli, in that they capture the important characteristics of English (or French) spelling.

1.5 The current state of play

Naming performance to letter strings has often been assumed to be "the least clouded window through which to examine the process of phonologic decoding" (Van Orden et al., 1990, p. 489). The above analysis has shown that the window needs cleaning. In a review, Reynolds and Besner (2005) attempted to do this; they identified three effects that they considered to be well-established in the literature, though not necessarily well-understood. These are (1) that a pseudohomophone advantage occurs in lists composed of both pseudohomophones and nonwords but with (2) no baseword frequency effect; and (3) a pseudohomophone with baseword frequency effect occurs in lists containing only pseudohomophones. However, as has been shown, the position is rather more complicated than this. The effect reported by McCann and Besner's was probably attributable to stimulus construction; Taft and Russell (1992) only achieved the effect for slow readers. Some studies have reported effects that were significant for by subjects but not by items (e.g. Borowsky, Owen & Masson, 2002, Expt. 1, and Herdman et al., 1996, Expt. 1, replicating McCann & Besner's items). According to Clark's classic paper, the fact that the by items analysis was not significant indicates that the results would not necessarily replicate with a new set of items (Clark, 1973).

In addition, in order to obtain the effect, some researchers have explicitly told participants the nature of the stimuli (e.g. Borowsky et al., 2002; Grainger et al., 2000; Marmurek & Kwantes, 1996). Since pseudohomophones have lexical status, and nonwords do not, it is hardly surprising to find that pseudohomophones are responded to differently in situations where participants are alerted to the nature of the stimuli. Such findings can tell us little about the nature of the word recognition system since a pseudohomophone effect emerging under such conditions might well occur because of post-lexical checking. Given that a pseudohomophone effect is so fragile, it is not surprising that a concomitant baseword frequency effect is also difficult to find; but such results were reported by Borowsky et al. (2002), Borowsky, Phillips, and Owen (2003), Marmurek and Kwantes (1996), and Grainger, Bouttevin, Truc, Bastien and Ziegler (2003). However, the co-occurrence of the two effects depended on the presentation order of the stimuli, that is, whether nonword or pseudohomophone blocks were presented first. Typically, when pseudohomophones were presented first, there was a frequency effect plus a pseudohomophone disadvantage; when nonwords were presented first, the effects weakened. However, this was not the case when Borowsky et al.'s used Seidenberg et al.'s stimuli, where the opposite results were reported.

So, despite Reynolds and Besner's optimistic attempt to establish a coherent state of affairs, it is really not clear how participants respond to pseudohomophones in naming experiments. Sometimes they are read faster than nonwords and sometimes not. Sometimes there is a co-occurring frequency effect and sometimes not. List composition and presentation order appear to affect results, but not in a consistent way. Although the literature offers a plethora of alternative theoretical interpretations, these interpretations are potentially insecure when they derive from such fragile data.

One possible reason for the inconclusive nature of the findings is to do with the experimental stimuli. As noted earlier, the use of nonwords is based on the assumption that spelling-to-sound mechanisms might be more easily uncovered by the use of these semantic-free stimuli. Stringent controls of one kind and another are then assumed to result in a set of items that are to all intents and purposes as similar to words as real, but unknown, words would be. However, inspection of the items used in typical naming experiments suggests that many stimuli are not like plausible new words; for example, *jinje* (McCann & Besner, 1987); *slej* and *bej* (Taft & Russell, 1992); *paije* and *soaz* (Borowsky et al., 2002). In practice, when a reader encounters a new word, this is more likely than not to be orthographically plausible. Monosyllabic neologisms that entered the 11th edition of the Concise Oxford

Dictionary included *radge, twonk* and *crunk*, which follow typical spelling patterns and are all clearly pronounceable; items that derive directly from foreign words, such as *mzee,* are fewer in number, and present a challenge as to pronunciation. While the underpinning assumption in the literature is that *shaip* is as similar to *shape* as *paiv* is to *pave,* one might equally well predict that different processes will be involved in processing the two pseudohomophones on the grounds that *shaip* is more plausible as a potential spelling than *paiv.*

It seems that controlling for N, bigram frequency, and so on, does not necessarily result in a letter string that is a plausible example of an English word. Some items are clearly visually more wordlike than others, and it is not unreasonable to suggest that readers are sensitive to orthotactic probabilities in the same way that they are sensitive to other aspects of language, such as grammar (e.g. Gerken & Bever, 1986) and phonotactics (e.g. Bailey & Hahn, 2001). Such sensitivity is a likely emergent property of a system in which frequently encountered spelling patterns develop stronger links with phonology than rarely-encountered patterns. It is difficult to see how the computational DRC model, with inbuilt lexical representations and a system of immutable GPC rules, could capture the concept of a nonword's visual wordlikeness, although it is possible that it might emerge via partial activation of real words in the orthographic lexicon (in the same way that N is thought to have its effect.) However, such an interpretation does not easily fall out of current theoretical or computational formulations of the DRC model; for example, Rastle & Brysbaert (2006) presented findings from priming experiments in the context of the DRC model based on a body of pseudohomophones and nonwords that contain some very unwordlike stimuli, such as *phloar* and *thiphe.*

1.6 Summary

If experimental data can demonstrate that readers process nonwords and pseudohomophones differently, this has important implications for models of reading; in particular, it can shed light on the status of lexical phonology. However, the current situation as regards naming experiments is confused, possibly as a result of the fact that most stimulus sets are heterogeneous in terms of how far items capture patterns of English orthotactics. The following chapters report a series of experiments, beginning with some naming studies and moving on to lexical decision and priming experiments, using stimuli expressly devised to be more wordlike, in order to explore the nature of the word recognition system, particularly in relation to the status of lexical and sublexical phonology.

Chapter 2

The effects of visual wordlikeness on pseudohomophone naming.

"It's a damn poor mind that can think of only one way to spell a word."
(Attributed to Andrew Jackson)

Although the position regarding pseudohomophone naming is very unclear, research at present indicates that pure lists elicit speeded naming over nonwords, sometimes together with a base word frequency effect (Borowsky et al., 2002; Marmurek & Kwantes, 1996, and Grainger et al., 2000); and such effects do not occur in mixed lists. Mixed lists on the other hand, are likely to elicit N effects. Overall these findings offer support for a view invoking the relative involvement of lexical and nonlexical routes in the different list conditions, and, as such, can be easily encompassed in a dual-route formulation. If such findings are indeed reliable, they are hard for the single-route account to explain convincingly. However, as the last chapter has indicated, the picture is not as clear as dual-route proponents would like to argue, and one reason for this may be to do with the construction of the experimental stimuli.

2.1 Stimulus construction in previous experiments

Studies that do not take account of nonword orthotactics open themselves up to serious problems of interpreting pseudohomophone and frequency effects. As indicated in the last chapter, many stimulus nonwords have simply not looked very much like English words. It could be argued, of course, that English is a heterogeneous language with influences from many different languages, so we should not be surprised by the strangeness of new items. One needs only to think of examples such as *raj, byte, chic,* and *feng shui* which have entered the language and which do not look like typical examples of English words. Normally, of course, when we come across such letter strings as new words, they would have some kind of semantic representation, but in the context of the research reviewed and reported here, the focus of interest is not on how reading operates when there is a semantic input (although maybe it should be). Rather, we are looking at what goes on when a reader comes across a novel word; and what we want to know is how the reader achieves a pronunciation for a typical group of letters. Until we know how readers deal with letter strings that are more or less typical of most of the items they are usually exposed to, it is not possible to deal with the atypical ones.

This is not to say that researchers in the past have strung letters together at random in order to create stimuli. Indeed, an attempt to achieve some kind of typicality has been embedded in attempts to control for various properties of words, so that controls have been made, at various times, for some or all of the following: frequency of base words; orthography in terms of summed positional bigram frequency and Coltheart N; onset and rime; length; number of syllables; legality of bodies. Furthermore, criticisms of stimulus creation have featured largely in theoretical discussions of the results reported in the literature, a trend that started after the initial work carried out by McCann and Besner (1987).

McCann and Besner's main criterion was the requirement that the base words for their pseudohomophones represented a wide range of frequencies. The only other controls were that the stimuli should be between four and six letters, with pseudohomophones and controls approximately matched on summed positional bigram frequency (SPBF). Their procedures resulted in some items that do not look like typical English words, such as *bawx, vawx, binje* and *jinje*. Marmurek and Kwantes (1996) employed McCann and Besner's procedure to create their nonwords, in spite of the criticisms that had been levelled against it (e.g. Taft & Russell, 1992). Using Monsell et al.'s (1992) high and low frequency lists, they took one- or two-syllable words and changed letters to make them into pseudohomophones, so that *tomb,* for example, became *toom*. A single letter was then changed to create a control nonword, e.g. *foom*. Again, this resulted in some peculiar-looking items, such as *lahf, tahf, playg,* and *stayg*.

Controlling for the orthographic factors they had identified as having been ignored by McCann and Besner, Taft and Russell (1992) tried to equate orthography across lexicality and frequency, by creating nonword triples with the same bodies (e.g. -*aim*), such that one nonword was homophonic with a high frequency word, one with a low frequency word, and one was non-homophonic (e.g. *gaim, daim, raim*). However, this procedure also produced some strange-looking triples, such as *milc, silc, filc; lej, slej, bej*. According to Herdman et al., (1996), Taft and Russell's high frequency items were more similar in terms of SPBF and N to their base words than the low frequency items, so that the frequency effect that Taft and Russell had claimed to detect was in fact an effect of orthography. Herdman et al.'s own new stimuli were created using a similar procedure to Taft and Russell in order to create triples, but they also equated high and low frequency items in terms of SPBF and N. They also created two types of orthographic rime status, that they termed 'legal' (where items ended in rimes such as *-ane* and *-air*) and 'illegal' (where items ended in non-existent orthographic rimes such as *-ayr* and *-urth*), in order to test the connectionist prediction that legal bodies should have stronger orthography-to-phonology mappings and therefore be processed more quickly than nonwords with illegal bodies. With these stimuli, it is noteworthy that the legal items tend to look more wordlike than the illegal items, many of which look decidedly odd, such as the triple *waije, raije* and *laije*. However, even among the legal items, some look more plausible than others; for example, *plawk* looks stranger than *prane,* possibly because *–awk* is a lower-frequency orthographic rime than *–ane* (frequencies of 3 and 13 respectively, in monosyllabic words). Therefore, it is possible that 'legality' per se is inadequate as a predictor of responses; and the relative frequency of orthographic rime components is an important consideration. Even then, distinguishing between items in this way does not capture the fact that the orthographic rime *–urth* is illegal and has zero frequency, but nevertheless produces relatively plausible nonwords such as *nurth* and *surth*. In short, it seems likely that the legal/illegal dichotomy may not be sensitive enough to capture all sources of sublexical information available to human readers.

Borowsky et al. (2002) carried out a series of pure and mixed list experiments using some earlier stimuli (Herdman et al., 1996; McCann & Besner, 1987; Seidenberg et al., 1996,), as well as their own new items. They controlled for base word frequency, length, and initial letters to create the pseudohomophones, and then changed one letter of the pseudohomophones to create the nonwords. However, the procedure again led to the construction of some items that do not look like plausible English words – such as pseudohomophones *dryv* and *mawths* and nonwords *bedj* and *mamths*.

Of all the stimuli created by previous researchers, those that intuitively look most "wordlike" are the ones created by Seidenberg et al. (1996). They created items specifically to control for problems they identified with McCann and Besner's stimuli. In particular, they pointed out that McCann and Besner's pseudohomophones were more wordlike than controls, since they began with more frequent onsets (such as *p, s, m* and *d*) than the nonwords which tended to begin with infrequent onsets (such as *kl, shr, z* and *ph*). Seidenberg et al. crossed pairs of onsets (e.g. *j* and *h)* and rimes (e.g. *–oak* and *–oap*) to form two pseudohomophones (*hoap* and *joak*) and two nonwords (*hoak* and *joap*). In terms of the protocols for creating stimuli described so far, this system seems to have the most advantages: in particular, it produces lists that are identical in terms of onset frequencies, bigram frequencies, and N. Nevertheless, it should still be noted that Seidenberg et al.'s stimuli still contain some 'odd' items such as the pairs *paije* and b*aije*, and *groaz* and *soaz*. This technique was also used by Reynolds and Besner (2005) to create a stimulus list including *phret*, *fiew* and *dwuild*. Choosing an alternative spelling to create the pseudohomophone seems to compromise the rime element in particular, with, for example, *–aije* chosen for *-age*, and *–oaz* for *-ose*, and this is potentially problematic. Herdman et al. (1996) identified naming differences for stimuli with existing and non-existent rimes, and Vanhoy and Van Orden (2001) established the salience of rimes in lexical decision experiments, establishing that pseudohomophones with extant rimes produced reliable lexical effects, while pseudohomophones with non-existent rimes did not. Therefore it is possible that unusual spellings, especially where the rime is concerned, elicit different response mechanisms or processes than those used for more commonly-encountered orthographic patterns.

2.2 Construction of new nonwords

Having considered the stimuli used by previous researchers, it is apparent that it is possible to make an intuitive judgement that some nonwords are visually more wordlike than others. Linguistic intuitions are widely used in the discipline of linguistics to provide evidence for theories (Devitt, 2006) but are less widely used in psycholinguistics. However, studies have reported judgements on various aspects of language: for example, co-reference (Gordon & Hendrick, 1997), grammaticality (e.g. Gerken & Bever, 1986), phonotactics (e.g. Bailey & Hahn, 2001), familiarity (Gernsbacher, 1984; Gilhooly & Logie, 1980), and frequency (Balota, Pilotti & Cortese, 2001). While these studies have reported judgements based on language users' implicit knowledge of their spoken language, it seems likely that readers should also be able to make judgements about their written language, and judge whether or not a letter string looks more or less plausible as an exemplar of English spelling. This is conceptualised as 'visual wordlikeness', and would seem to be an orthographic variable, separate from traditional measures of SPBF and N, because, as has been seen, controlling for these measures does not always prevent 'strange' items from emerging in the stimulus lists. Knowledge of orthographic rime might well be a factor, but there may be other orthotactic contributions as well; such as the relative frequency of various letter combinations in all parts of the letter string.

The studies reported below started by the creation of a stimulus set that in theory should be more wordlike than previous sets; the new items were rated for wordlikeness. The subsequent experiments then investigated whether the creation of a more wordlike set of stimuli could provide clearer information about phonological factors in naming and lexical decision experiments. The experiments also enabled an investigation of whether participants responded to the wordlikeness of the stimuli.

The procedure adopted by Seidenberg et al. and Reynolds and Besner in swapping onsets and orthographic rimes provides a useful protocol because it leads to homogeneity between lists of pseudohomophones and nonwords. However, one of the problems has been the fairly arbitrary method of creating the pseudohomophone spelling and this appears to have led to some of the problematic items. Starting with real words gets around this problem, and English is ideally suited to this approach because alternative orthographic representations frequently exist for the same phonology. In particular, multiple spellings exist for rimes, and by taking pairs of "feedback inconsistent" words (such as *soap* and *pope*) it is possible to construct pseudohomophones that are entirely created from existing words (in this case, *sope* and *poap*). Since around 72% of English monosyllables are feedback inconsistent (Ziegler, Montant, & Jacobs, 1997), this technique is likely to produce a stimulus set where all the items are very representative of the population of monosyllabic words.

2.2.1 Method

Participants

16 students at the University of Bristol carried out a pen-and-paper pilot study. 65 different students gave wordlikeness ratings for pseudohomophones, and 48 more students gave ratings for the nonwords. Participants in this and all the following experiments were non-dyslexic, had English as their first language and had normal or corrected-to-normal vision.

Stimulus materials and procedure

Pseudohomophones were created by swapping onsets and rimes of pairs of words, as described above, from a set of 1,683 English monosyllabic words taken from the CELEX database (Baayen, Piepenbrock & van Rijn, 1993) (including monosyllabic words with CVC structure, and excluding inflected forms and where there were at least two alternative spellings of the phonological rime, e.g. *chalk, hawk, cork*). This technique produced a large number of such pairs with a wide variety of legal rimes and unambiguous pronunciations. From the larger list, 99 pseudohomophones with a wide variety of onsets, rimes, and base word frequency were chosen. In order to check, first, that people could actually rate these items in terms of wordlikeness, and, second, that their ratings would be reliable, a pen-and-paper pilot study was carried out. 16 students were asked to look at the 99 items and rate them on a scale from 1 – 7, according to how much they looked as though they could be real English words. Participants in this pilot study were explicitly asked to try to ignore any pronunciation generated as they silently read the letter string. Agreement among participants was good (all with a correlation of more than 0.5 with the overall mean), apart from one participant, whose data were excluded.

Standardised mean ratings (z scores) were calculated for each item in order to compensate for differential use of the rating scale between raters. 52 items were chosen as the critical stimuli, selected on the basis of having the highest level of agreement among participants in the pilot study. Each item had a different rime, although there was some repetition of onsets. Log_{10} base word frequency for the items ranged from 0 to 4.34 (raw frequency 1 - 21,740), and the items were normally distributed within this range. N ranged from 0 to 14; 22 of the items were neighbours of their base word. The new stimuli were then mixed with the pseudohomophones used in previous research and presented to a new group of 65 students as a computer-based task. Each participant rated the visual wordlikeness of a quasi-random sample of 98 items. Individual ratings were correlated with the grand means for all items; all correlations were $> .5$, indicating a high level of agreement, except for 2 participants whose data were excluded. Correlations of the wordlikeness ratings for the 52 items between this and the pilot study gave $r(52) = .72$, $p < .0001$, suggesting that the task is reliable. Weak wordlikeness correlations with N ($r(52) = .26$, $p = 0.03$) and base word frequency ($r(52) = -.21$, n.s.) suggested that raters may have been influenced by these two factors, so their contributions were partialled out of the ratings to give a set of standardised residual z scores that were used in all the analyses reported here.

To create the nonwords, the onsets and rimes of the pseudohomophones were swapped; for example, pseudohomophones *poap* and *clame* produced *cloap* and *pame*. A small number of onsets and rimes were changed in order to create viable nonwords, but the final lists of 52 pseudohomophones and 52 nonwords were substantially identical in terms of onsets and rimes. The nonwords were rated, together with those used in previous research, in the same way as the pseudohomophones, by a new group of participants. (See Appendix A for stimuli and wordlikeness ratings).

2.3 Experiment 1:
Naming pseudohomophones and nonwords

Aims and predictions

The aim of the first experiment was first, to see if pseudohomophones and nonwords created from existing monosyllabic words could establish a more systematic pattern of results than that seen in previous studies, and second, to see if visual wordlikeness had an effect on pseudohomophone and nonword naming. More specific predictions are that:

- A pseudohomophone effect with a co-occurring baseword frequency effect will be more apparent in pure than in mixed lists.
- List presentation order will affect pseudohomophone naming.
- Wordlikeness effects are more likely in mixed than in pure lists.

Following Experiment 1, computer simulations were carried out, using the new stimuli, which were presented to the DRC computer model (Coltheart et al., 2001) and two versions of the models reported in Plaut et al. (1996). Given that the connectionist models in general learn from exposure to statistical properties inherent in spelling patterns and in particular the Plaut et al. models are based on onset-rime input coding properties, it might be expected that these models will show an effect of wordlikeness. For the DRC model, a wordlikeness effect might emerge from the grapheme-to-phoneme route if wordlikeness captures the same rules as those implemented in the model.

2.3.1 Method

Participants

45 University of Bristol students participated in the experiment for partial course credit. Participants were assigned at random to list conditions so that 15 students read pseudohomophones first, followed by control nonwords, 15 read control nonwords first, then pseudohomophones, and 15 read mixed lists.

Apparatus

Presentation of stimuli and recording of spoken responses were controlled by customised software created by Dr. Clive Frankish in the Bristol University labs, running on a PC with an Iiyama Vision Master Pro 413 CRT monitor. Target letter strings appeared as black lower case characters in Courier New font, size 50pt on screen, on a grey background. Participants viewed the monitor screen from a distance of 60 cm, and the entire letter display subtended a visual angle of 11° x 5°. Participants' spoken responses were recorded via a microphone headset. The experimental session was broken into four blocks of 52 trials, with rest pauses between blocks. Within these blocks, trials were run in a continuous sequence at a rate of 3 sec per trial. Each trial began with a display of a fixation cross in the centre of the screen for 250 ms, followed by the target display, which remained visible for 800 ms. Digital recording was initiated when the target display appeared, and continued for a duration of two sec. These recordings were subsequently analysed using speech editing software to determine naming latencies.

Stimuli and design

The stimuli were generated as detailed above. All were monosyllabic nonwords of four or five letters, with legal onsets and rimes. Lists of two types, pure and mixed, were created, using filler items to ensure an equal number of items in each list. Pure lists contained either 52 critical pseudohomophones and 52 filler pseudohomophones, or 52 critical non-pseudohomophonic nonwords and 52 filler nonwords. Stimuli were presented to the participants in quasi-random order. Participants named either two pure or two mixed lists, and presentation order was counterbalanced so that half the participants in the pure lists condition named pseudohomophones first, control nonwords second, and the other half of the participants named control nonwords first, pseudohomophones second. Presentation of the mixed lists was similarly counterbalanced. A pen-and-paper lexical decision task was also created, containing a randomised list of all the critical pseudohomophones and nonwords.

Procedure

Participants were assigned to one of two conditions on the basis of an alternating sequence. If they were assigned to pure lists, they were either given the pseudohomophone list first and the control nonword list second, or the controls first and the pseudohomophones second. If they were assigned to the mixed lists, they were given either the critical items first and the filler items second, or the fillers first and the critical items second. All participants therefore named 208 items, which took approximately 20 minutes.

Participants were told, orally and in writing, that they would see a sequence of pronounceable letter strings which they were asked to pronounce as quickly and as accurately as they could. They were not told the nature of the letter strings. Each trial was preceded by a practice trial of 10 items, either pseudohomophones, control nonwords, or mixed items, in order to familiarise participants with the task and also to nudge them into using appropriate strategies to generate pronunciations for the type of list with which they would be presented.

Participants sat in front of a computer at a comfortable distance from the screen in a dimly lit, quiet room. Stimuli, in white, appeared singly in lower case letters in the centre of a dark grey screen. Participants initiated sessions by pressing the space bar and a fixation cross appeared in the centre of the screen followed by the letter string. Participants were offered the opportunity to rest after every 26 items; if they did not need to take a break they pressed the space bar to continue.

After completing the trials, participants immediately completed the phonological lexical decision task. This was a pen-and-paper exercise, and they were asked to tick those items they considered would sound like a real word if spoken aloud. Participants had not been made aware of this task before they engaged in the naming task. Participants were then debriefed as to the purpose of the experiment.

The recordings of the participants' responses were replayed by the experimenter, who inspected responses on an item-by-item basis. Each item was displayed as a visual trace and re-played against the audio recording, in order to check that the algorithm embedded in the customised speech editing software (developed in the Bristol University labs, as above) had correctly detected the onset response latency. Errors of pronunciation were also identified during this process, and omitted from the data analysis.

2.3.2 Results

Overall mean latencies and standard deviations for correct responses to the critical items are shown in Table 1. Analyses were performed on trimmed data; scores more than 2.5 standard deviations below or above an item mean were classed as outliers and excluded. In this and all subsequent analyses, an alpha level of .05 was used for all statistical tests.

Table 1
Means, standard deviations and errors for pseudohomophone naming in pure and mixed lists.

	Pseudohomophones		Nonwords		
	Means*(S.d.)* (ms)	Errors (%)	Means*(S.d.)* (ms)	Errors (%)	Difference
Overall mean	518 *(40)*	5.7	529 *(42)*	7.5	-11
Mixed lists overall	514 *(41)*	4.0	518 *(40)*	6.9	-4
Pure lists overall	521 *(39)*	7.4	539 *(45)*	8.1	-18

The data were initially analysed to see if pseudohomophones were named faster than nonwords, and if list type and presentation order affected pseudohomophone naming latency. A mixed 2 x 2 analysis of variance was carried out with list type (pure or mixed) as the between subjects factor, and nonword type (pseudohomophone or true nonword) as the within subjects factor. By items analysis showed that list type was marginally significant ($F(1,102) = 3.95$, $p = .05$) while nonword type was significant ($F(1,102) = 7.01$, $p = .01$). The interaction between list type and nonword type was not significant ($F(1,102) = 3.36$, $p < 1$). By subjects analysis showed that list type was not significant ($F(1, 43) = .27$, n.s.) while nonword type was not significant ($F(1, 43) = 3.27$, $p < 1$). The interaction between list type and nonword type was not significant ($F(1, 43) = 1.76$, n.s.).

Planned comparisons were carried out to compare pseudohomophones with nonwords. In this, and all subsequent analyses, independent samples t tests were used for the by-items analysis, and paired samples t tests for the by-subjects analysis; all tests were 2-tailed. Pseudohomophones were not read faster than nonwords in the by items analysis, t2 (102) = -1.44, n.s.; although the by subjects analysis was significant (t1 (44) = -2.37, p = .022). Mean latencies indicate that, in pure lists, pseudohomophones appear to be read faster than nonwords, but the difference is clearly smaller in the mixed lists (see also Figure 4). The difference between pseudohomophones and nonwords was significant in pure lists (t1 (29) = 2.26, p = .03, t2 (102) = 2.27, p= .03), but not in mixed lists (t1 (14) = .94, n.s.; t2 (102) = .43, n.s.).

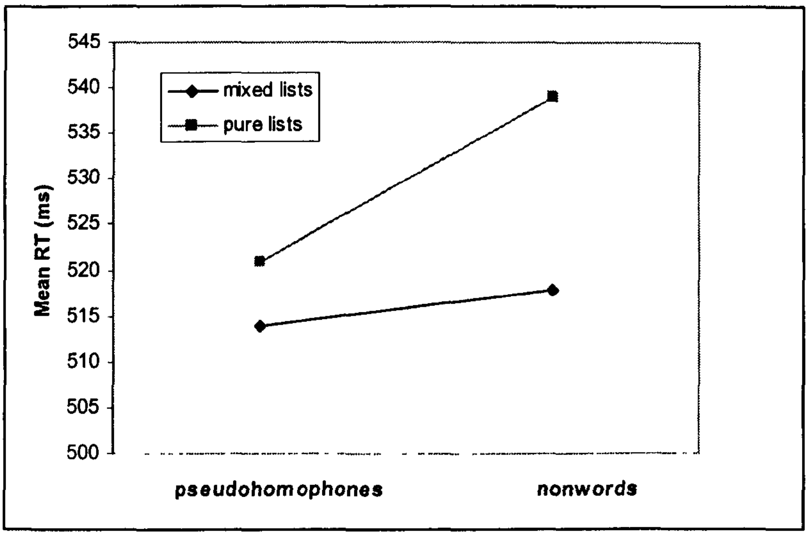


Figure 4. Pseudohomophone and nonword latencies by list type (pure or mixed).

Table 2
Reaction times in ms for pseudohomophone and nonword naming as a function of list presentation order.

		Pseudohomophones	Nonwords	Difference
		Mean RT *(S.d.)*	Mean RT *(S.d.)*	
Mixed lists				
	PHs first	533 *(47)*	535 *(46)*	2
	PHs second	478 *(42)*	486 *(41)*	8
Pure lists				
	PHs first	545 *(49)*	537 *(48)*	8
	PHs second	497 *(36)*	541 *(45)*	-42

List presentation order

Table 2 shows that pseudohomophones appear to have been named faster than nonwords when they were presented in a pure list after nonwords, but when they were presented first, there was very little difference in naming times. This effect of order presentation does not seem to be apparent with the nonwords, where the latencies are essentially the same regardless of list presentation order (see also Figure 5). To examine the difference in pure lists, t tests were carried out. When pseudohomophones were read first, there was no significant difference between naming latencies for pseudohomophones and pure nonwords ($t1\ (29) = 1.07$, n.s.; $t2\ (102) = .71$, n.s). However, when pseudohomophones were read second, they were significantly faster than nonwords ($t1\ (29) = 4.06$, $p = .004$; $t2\ (102) = 5.26$, $p < .001$).

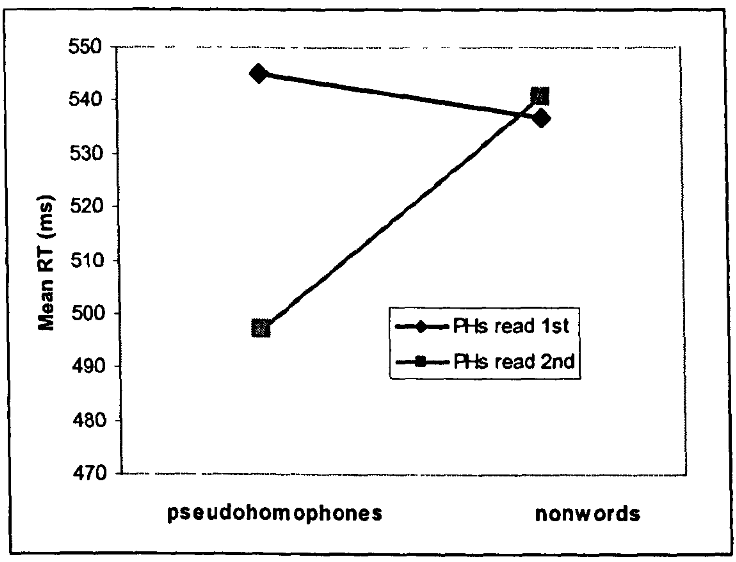


Figure 5. Pseudohomophone and nonword naming latencies by presentation order of pure lists.

To explore whether pseudohomophones were selectively read faster when they were presented in a second block, regardless of list composition, mean reaction times for pseudohomophones in mixed lists was also analysed by presentation order. Figure 6 shows that pseudohomophones were not selectively advantaged over nonwords when read in a second block, so the speeded naming is not attributable to a selective advantage that somehow facilitated pseudohomophones in a second block irrespective of list type.

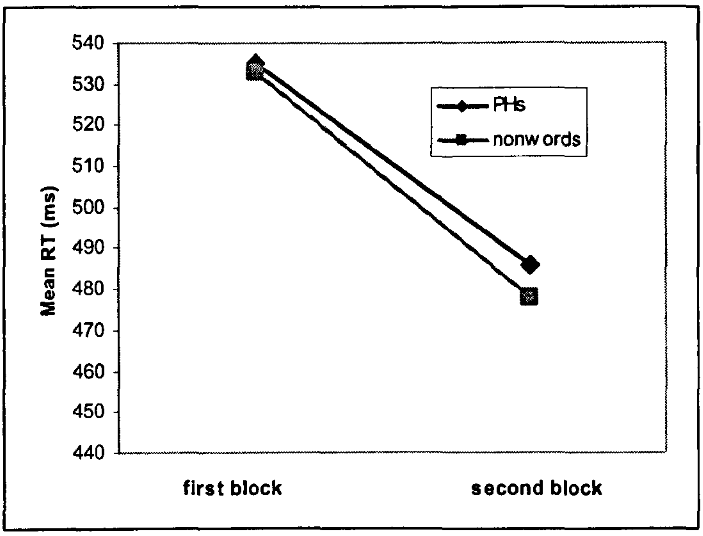


Figure 6. Pseudohomophones and nonwords in mixed lists both speed up when presented in second block.

So far, the only evidence for a pseudohomophone effect is in a pure list read after a list of nonwords. To explore whether this could be because of a within-list practice effect that only occurred when pseudohomophones were read second, the lists were divided into four, and the mean latency for each quarter established separately. The pattern of reaction times to pseudohomophones is remarkably similar; the only difference is that when they are presented second, they are read faster (Figure 7).

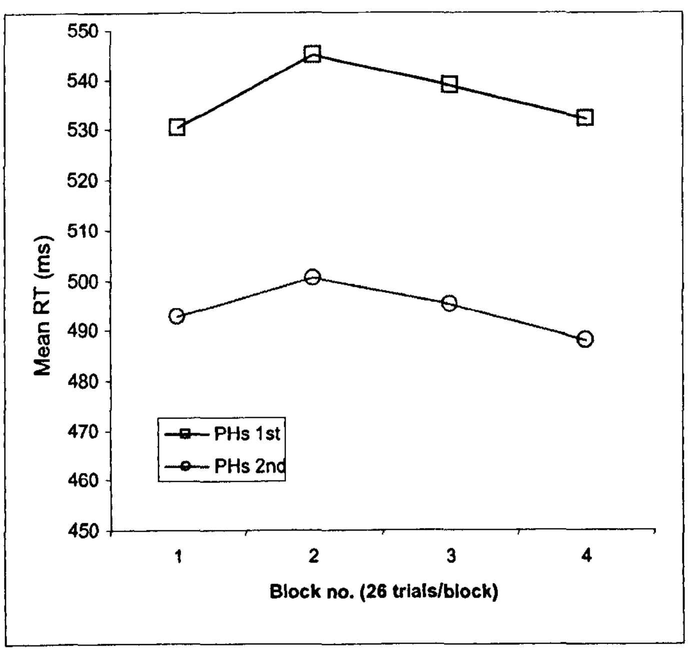


Figure 7. Pseudohomophone naming times by list sub-sections offer no evidence for a within-list practice effect when pseudohomophones are read second.

Effect of list composition on frequency, wordlikeness and N

If participants are making selective use of a lexical and a nonlexical route in naming items in different list types, then we would expect to see different effects for baseword frequency, N and wordlikeness. Based on previous research, we might expect a baseword frequency effect to emerge in pure lists, in tandem with a pseudohomophone effect, because lexical representations are activated. Wordlikeness, as a measure of orthotactic knowledge, is more likely to emerge when the nonlexical route is operational, as in mixed list reading. An effect of N is more difficult to predict – it has been found to affect word, nonword and pseudohomophone naming in both types of list (Grainger et al., 2000). Therefore, pseudohomophone naming latencies in mixed and pure lists were correlated with these measures. The frequency measure was $\log_{10}$, the N measure was the number of neighbours where each neighbour word had a CELEX frequency of 10 or more (corresponding to approximately 0.6 occurrences per million), and the wordlikeness ratings were the standardised residual scores with contributions from N and frequency partialled out, calculated as described earlier.

Table 3

Correlations between pseudohomophone naming latency in pure and mixed lists with base word frequency, N, and wordlikeness.

	Log_{10} freq	N (10)	Wordlikeness
Mixed lists	-.21	-.30*	-.42**
Pure lists, PHs first	-.24	-.18	-.32*
Pure lists, PHs second	-.23	-.03	-.24

** = p <.05, ** = p < .001*

Table 3 shows that N correlated with latencies in mixed lists, but not in pure lists. Wordlikeness correlated significantly in mixed lists, and in pure lists but only when pseudohomophones were read first. This might be consistent with a shift between lexical and nonlexical or sublexical strategies; when pseudohomophones are read mixed with nonwords, or when they are read first, wordlikeness assumes greater importance. However, the traditional marker of lexical access, baseword frequency, does not give any support to this view; the pseudohomophone effect only occurred in the naming second condition but baseword frequency correlated no more strongly in this condition than in any other.

It would be possible at this point to carry out a stepwise regression analysis to investigate the roles of frequency, N and wordlikeness as predictors to latencies in pure and mixed lists, but such analysis would be redundant. Since frequency and N are uncorrelated with each other for these stimuli ($r\ (52) = 0.07$, $p = .64$), and their contributions have been partialled out of the wordlikeness ratings, all three variables are uncorrelated. Therefore, the relative roles of the variables of interest are indicated in Table 3.

Individual differences in reader speed and skill

Since previous work (e.g. Laxon et al., 1992; Seidenberg et al., 1996; Taft & Russell, 1992) has suggested that an otherwise undetectable pseudohomophone effect might be apparent in readers labelled as less skilled, less accurate, or slower, an analysis based on subject speed was carried out. The mean speed for all 45 participants in both conditions together was calculated, on the basis of their response latencies to pseudohomophones and nonwords. There was no significant difference in mean naming speed depending on which condition was undertaken ($F(2) = .406$, n.s.), so it is justifiable to treat them as a homogeneous group in order to carry out an initial analysis.

The reading of pseudohomophones and nonwords was compared for the fastest and slowest 15 participants in all conditions together. Although a marginally significant pseudohomophone effect was detectable, it was present for the fastest readers ($t1\ (14) = 2.1$, $p = .054$, $t2\ (102) = 1.76$, $p = .08$). Reaction times correlated with baseword frequency ($r\ (52) = -.32$, $p = .02$). There was no difference in latencies to pseudohomophones and nonwords for the slowest readers ($p > .6$ in all cases) and there was no frequency effect. But there was a significant correlation with wordlikeness for these participants ($r\ (52) = -.36$, $p = .008$), and a weaker correlation with N ($r\ (52) = .28$, $p = .048$).

Controlling for items rated as pseudohomophonic

Borowsky et al. (2002) suggested that only those items rated as pseudohomophonic by participants should be analysed. Removing data associated with all the items that participants did not recognise as sounding like real words (as determined by the post-experiment lexical decision task) did not alter the overall pattern of results as reported above. An unanticipated side-effect of this procedure meant that 16 of the 45 participants now scored below 75% correct responses on their naming latencies; when these participants' data were removed from the analysis entirely, there was again no change to the overall pattern of results.

Error analysis

Errors were totalled for each pseudohomophone; they were identified from three procedures already mentioned. First were errors of pronunciation identified by the experimenter; second were errors identified during the reaction time trimming procedure; and third were those included because the participant did not recognise them as sounding like a real word. Correlations between number of errors and the measures of frequency and N were not significant, but the correlation with wordlikeness was (r (52) = -.316, p = .023), indicating that the more wordlike an item was, the less likely it was to generate errors.

Summary

To summarise; pseudohomophones were named faster than nonwords, but the effect was moderated by a number of factors, including list composition; whether the pure pseudohomophone list was read before or after the nonword list; and subject speed or skill. As expected, the orthographic measure N was significant in mixed lists, although contrary to predictions, a base word frequency effect did not co-occur with the pseudohomophone effects. The only co-occurrence of these two variables was demonstrated by the fastest readers. It was thought that wordlikeness would be of importance in mixed rather than pure lists, but it showed significant correlations with latencies in both pure and mixed lists, although the effect was stronger in mixed. Wordlikeness was also shown to be a factor in the number of errors for each item.

2.3.3 Naming by computer models

A simulation of Experiment 1 was carried out using the DRC model (Coltheart et al., 2001; http://www.maccs.mq.edu.au/~max/DRC/) to 'name' the 52 critical pseudohomophones and 52 nonwords. Pseudohomophones were read faster than nonwords (t (51) = 2.99, p = .002), but the number of cycles to name pseudohomophones did not correlate with the experimental data in either the pure or the mixed lists (r (52) = -.10, n.s., pure lists; r (52) = .04, n.s., mixed lists).

Correlations with pseudohomophone base word frequency, N and wordlikeness showed that only N gave a significant correlation. As shown in Table 4, the weak positive correlation with wordlikeness was in the *opposite* direction to the one shown by human readers: the model took longer to 'read' more wordlike items.

Table 4
Correlations between DRC cycles to naming pseudohomophones with base word frequency, N and wordlikeness.

	Base word freq	N	WL
DRC naming	r = -.18	r = -.32*	r = .23, *p* = *.1*

** = p <.05, ** = p < .001*

An additional simulation using the experimental stimuli was carried out using two versions of the connectionist models reported in Plaut et al., (1996)[1]. The first of these used the attractor network reported in Simulation 3; and the second used the feedforward network with additional input representing semantics reported in Simulation 4. The attractor model read pseudohomophones more slowly than nonwords (t (72.2) = 2.1, p = .04), but the feedforward model did not (t (101) = 1.17, n.s.). There is no theoretical reason why the attractor model should settle faster for pseudohomophones than nonwords, since the network's substructure reflects only sublexical, not lexical, correspondences. Inspection of individual items revealed that the model produced particularly long settling times for *poze*, *churp* and *fome*, and removal of these stimuli abolished the apparent pseudohomophone effect. Neither model showed significant correlations with the naming latencies, but there was a

[1] I am very grateful to Professor David Plaut for carrying out these simulations.

weak correlation with wordlikeness and the attractor model's settling times (see Table 5). Overall the qualitative fit to the human data is slightly better than the results produced by the DRC simulation.

Table 5
Correlations between connectionist networks' performance, participants' naming latencies, and base word frequency, neighbourhood and wordlikeness.

	RTs mixed	RTs pure	Frequency	N	WL
Attractor model Simulation 3 (n = 49)	.12	.20	-.22	.01	-.28 *p < .1*
Feedforward model Simulation 4 (n = 52)	.11	.15	-.08	-.19	-.20

* = *p* <.*05*, ** = *p* < .*001*

2.3.4 Discussion

The aim of Experiment 1 was to see if a stable pattern of pseudohomophone and base word frequency effects would emerge when nonword stimuli were created according to rules governing English spelling. A second aim was to see whether participants would show sensitivity to the wordlikeness of these stimuli, which would lend support to the argument that results of previous research may be difficult to interpret because of the nature of the stimuli used. In Experiment 1 we have seen a significant pseudohomophone effect in pure lists but without a baseword frequency effect; and an apparent dissociation between fast and slow participants in that the fastest participants showed co-occurring pseudohomophone and baseword frequency effects while the slowest readers showed effects of N and wordlikeness. Because a baseword frequency effect occurs in one circumstance and not the other, it is probable that the pseudohomophone effect emerges from two different sources. The pure list finding can be explained in terms of readers making use of appropriate strategies, and setting a suitable response-time criterion, while the reader speed finding can be explained in terms of skilled readers making use of their highly-interconnected system of knowledge about lexical and sublexical spelling-to-sound correspondences. The former is not very informative about how reading is normally accomplished, whereas the latter can add to our conceptualisation of the underpinning reading system.

Pseudohomophones in pure lists

In the overall analysis, there was only a single instance of the pseudohomophone effect: when a pure list was read after a list of control nonwords. When pseudohomophones were read before nonwords, or when they were read in mixed lists, they were not read faster than nonwords. The mixed list findings of a null pseudohomophone effect replicate previous work (e.g. Grainger et al., 2000; Marmurek & Kwantes, 2006). It could be argued that reading a second list of any kind will speed latencies, but this cannot be the case, or nonword latencies would also be speeded when they are read second, and they are not (Table 2). It might also be argued that pseudohomophones are, for whatever reason, selectively read faster than nonwords when they are read second, regardless of list type; but they are not (Figure 6). It could also be argued that there is a within-list practice effect that selectively facilitates pseudohomophones when they are read after a list of true nonwords, but this also is an untenable position (Figure 7). So why are pseudohomophones selectively advantaged when read second? Given that pseudohomophones and nonwords were constructed to be as identical as possible, it cannot be argued that the pseudohomophones are different from the nonwords in anything other than the fact that they sound like real words when spoken aloud.

There are two possible explanations. The first is that pure pseudohomophone lists are more likely to elicit lexical rather than sublexical processing, since all items in such lists have lexical status; half of the items in mixed lists do not have lexical representations, so the emphasis is more likely to be on sublexical processing. Various studies have shown that readers are sensitive to list composition effects in terms of relative activation of lexical and sublexical knowledge (e.g. Content & Peereman, 1992; Monsell, et al., 1992) and can shift attentional control depending on list demands (e.g. Reynolds & Besner, 2005; Zevin & Balota, 2000). However, if pseudohomophones were accessing base word representations in pure lists, then a concomitant frequency effect would be expected, but this was not in evidence. Further, the effect should be the same irrespective of whether the list was read before or after a pure list containing true nonwords. We might accept that experience with a list of nonwords had the effect of alerting participants to the lexical status of the second list and therefore they placed more reliance on processes associated with lexical phonology, rather than the sublexical processes used for the first list, so the effect only emerged in this condition. We would have to then accept that lexical effects were occurring without a base word frequency effect. If participants are accessing base words, we would have to assume that the phonological lexicon is not frequency-sensitive (as was argued by McCann & Besner, 1987).

Alternatively, if participants become aware that the second list, by comparison with the first, generated a familiar output, this might well speed response times. This is a feedback effect, occurring after spelling-to-sound has been computed, and might occur at articulation, or alternatively be located as a result of auditory feedback from speech output (Price, Moore & Frackowiak, 1996). This explanation does not implicate a base word frequency effect. Evidence for this approach was provided by Lupker et al. (1997) and Kinoshita and Lupker (2002, 2003), who suggested that there is a "time-criterion effect" on the time taken to start articulation, dependent on whether the list contains homogeneous or heterogeneous stimuli. Participants set a naming criterion appropriate for the task they are doing, setting the fastest possible time that can be achieved without running the risk of making an error. Mixed blocks tend to speed up responses that would, in pure blocks, be slower, and slow down responses that, in pure blocks, would be faster. Table 1 shows that the slowest pure block nonword naming speed was 541 ms, and the fastest pure block pseudohomophone time was 497 ms; together giving a mean speed of 519 ms, which is remarkably close to the average mixed block naming speed of 516 ms. This explanation differs from the first in that it is based on the idea that participants set a response deadline appropriate to the task in hand; the first explanation suggests that participants selectively attend to different routes or processes depending on the task in hand.

The two explanations are not necessarily mutually exclusive, but in this case, since pseudohomophones were only read faster when they were read after a list of nonwords, the response-setting criterion seems more likely than the attentional control hypothesis, which does not address list order effects. When read first, orthographically wordlike pseudohomophones were read at the same speed as the apparent default speed for true nonwords i.e. around 540 ms, so it seems that the reading system in this case treated both sets of stimuli in the same way. Evidence that participants in this experiment were treating pseudohomophones when read first more like true nonwords, is shown by the fact that the correlations for N and the wordlikeness variables were stronger for the pseudohomophones first than for the pseudohomophones second list. Overall, therefore, the speeded pseudohomophone naming in the second list is better explained by the response-setting criterion approach rather than the attentional control hypothesis. This is important because it occurs as a result of cognitive mechanisms that are separate from those involved in the reading system per se.

Finally, we need to explain why previous research has tended to find a pseudohomophone *dis*advantage with an accompanying base word frequency effect when pseudohomophones are read first. Data in these cases has resulted from experiments where researchers have told participants about the pseudohomophonic nature of the stimuli (e.g. Borowsky et al., 2002; Grainger et al., 2000; Marmurek & Kwantes, 1996). Under these conditions, it seems more than likely that participants would consciously adopt lexical strategies, and actively seek to identify the base word. This would lead to slower naming and a co-occurring base word frequency effect. This view that participants were engaging in extra processing resulting from experimenter instructions is supported by the fact that latencies for all four sets of stimuli used by Borowsky et al. are around 150 ms longer than those reported in Experiment 1; and for the monosyllables used by Marmurek and Kwantes (1996), the increase in latencies is around 250 ms.

Are the results caused by flawed stimuli or methodology?

Overall, the evidence so far suggests that the pure list pseudohomophone effect in Experiment 1 occurred simply because participants set a faster response time criterion for producing familiar-sounding items. It is possible, of course, that the results reported above arose from flawed stimuli or methodology. In terms of stimuli, it was an explicit aim of Experiment 1 to control the creation of nonwords so that both types of nonword were plausible, legal exemplars of possible English words. In particular there were no illegal orthographic rimes; rimes were capable of only one pronunciation (in theory, at least; in practice, readers sometimes generated creative alternatives); and onsets of pseudohomophones and nonwords were common and well-matched. Therefore, if the stimuli were confounded in some unknown way, it is difficult to see why this advantaged one group of readers over the rest. As far as the method is concerned, the experiment aimed to replicate those used in recent work with the use of pure and mixed lists, (e.g. Borowsky et al., 2002, Marmurek & Kwantes, 1996). The only difference in procedure was that participants in Experiment 1 were not told that some of the items were pseudohomophonic, for the reasons given above. Analysing only those items recognised as pseudohomophonic, as recommended by Borowsky et al., did not result in any changes in the overall results. It is difficult therefore to argue that the methodology employed meant that an otherwise detectable and genuine pseudohomophone effect, possibly linked to baseword frequency, was obscured by some effect of the experimental procedure. Instead, it seems clear that a pseudohomophone effect is not an inevitable outcome for most readers when naming nonwords, as long as those nonwords are wordlike. The fact that pseudohomophones and nonwords are both read using the same processes has interesting implications for models of reading. Where pseudohomophone effects have been reported in experiments using less wordlike stimuli, they may have emerged as a result of problem-solving strategies employed to deal with unfamiliar orthotactic patterns, which result in different processing times for pseudohomophones and nonwords. Thus, the suggestion is that there is no difference in naming times for *brane* and *brate*; but there may be a difference between *brayn* and *brayt*. It might be that the additional time needed to process less wordlike items allows information about phonological lexical status to arise, and to affect processing. However, information about processing times for the more wordlike items is arguably of more relevance in furthering our understanding of the reading system, since they are more representative of English orthography than the unwordlike items.

Theoretical and computational models

Dual-route models

In general, Experiment 1 did not provide evidence for pseudohomophone and frequency effects (the significant findings for fast readers are discussed in the next section). The general findings can be interpreted in terms of a fast and efficient assembly procedure that produces an output phonology for orthographic input without the need for additional input from lexical representations generated via a second route. Whether inputs are pseudohomophones or not is irrelevant when the input adheres to the regularities of English spelling; pronunciations can be generated without additional processing from lexical representations. In this case, it would be expected that nonwords and pseudohomophones would be named in the same times, and this is shown in the data when participants name either sort of stimulus first (i.e. are not employing a strategy based on prior experience within the experimental session). The connectionist account encompasses this explanation neatly because it makes precisely this kind of assumption; *soke* and *sote* are processed in exactly the same way.

Theoretically the data may also be explained in terms of the lexical and nonlexical route of the dual-route model, but the fit between the human data and the DRC's computational output was not convincing. The computational model gave a pseudohomophone effect, a weak wordlikeness effect, and a stronger N effect for the experimental stimuli; such effects are predicted by the model, but this is not what the human data showed. It may be that altering the DRC parameter-settings would enable a better simulation of the human data, but this is by no means certain. Apart from anything else, the model responded to the more wordlike items more slowly than to the unwordlike items, the opposite pattern to the experimental data. If the model's inbuilt grapheme-to-phoneme rules cannot generate responses that reflect human sensitivity to wordlikeness, then altering its parameters will probably not make much difference to its behaviour.

A connectionist dual-route approach such as that outlined by Zorzi et al. (1998) could also account for the findings. The Two Layer Assembly (TLA) model proposes that nonwords would be read via the nonlexical feedforward route; this route has identified sublexical spelling-to-sound mappings from its training sessions, and can generalise this knowledge to nonwords. Whether or not stimuli are pseudohomophonic is irrelevant. Contrary to the assumptions embedded in standard-dual route models, this route operates more quickly than the hidden, quasi-lexical, pathway, so even if this second pathway can make a contribution to output, it will come too late to have an effect. Although it was not possible to carry out a simulation with the Zorzi et al. model, it does seem that a model with a fast nonlexical route that has learned about the statistical properties of words is in theory better able to explain the results of Experiment 1 than a model with a slow nonlexical route with in-built rules.

Single route models

Proponents of single-route models state explicitly that there are no lexical representations in the system and that where pseudohomophone and base word frequency effects have been established in the past, it is as a result of post-assembly processes. Seidenberg et al. (1996) explain pseudohomophone and, where they occur, frequency effects, in terms of motor articulation; word sounds are produced more frequently than nonword sounds, and this facilitates naming. Since Experiment 1 has shown clearly that there is no pseudohomophone or frequency effect, this might be taken as support also for connectionist models. Inputs that are samples of the stimuli that the network has learned on are processed according to the regularities discovered by the network. The computational simulations carried out using the Plaut et al. (1996) computational models offer some additional evidence for this view. As with the data from Experiment 1, neither model gave a pseudohomophone advantage over nonwords, and neither model showed a base word frequency effect. In the attractor model, wordlikeness was marginally significant ($p = .057$) and the correlation was in the same direction as the human data; this suggests that the model during training has developed a sensitivity to some of the same spelling-to-sound statistical regularities that human readers also demonstrated. There was no correlation with N. In broad terms, the results given by the connectionist models are rather closer to the human data than those from the DRC model.

Reader speed

One way in which we might explore the claims of competing models further is to look at the finding that fast readers demonstrated clear pseudohomophone and frequency effects, irrespective of which presentation condition they were in. These effects are at odds with any research that has investigated reader speed; Seidenberg et al. (1996) found no effects across fast, medium and slow readers, while Taft and Russell (1992) found the effects for slower readers only. (It is likely that the Taft and Russell findings are not reliable; they based their criterion for reading on just nine nonwords, and the frequency effect was only significant for some of the stimuli. Moreover, Marmurek and Kwantes (1996) failed to replicate their reader speed findings). We have seen that the slowest readers in Experiment 1 did not show a pseudohomophone or a base word frequency effect – but they did show significant effects of N and wordlikeness. It appears that the fastest readers accessed or activated lexical phonology in a way that did not characterise the slowest readers, who in turn were more reliant on, or sensitive to, the orthographic measures of N and wordlikeness. It seems unlikely that these novel findings are attributable to some undetected confound in the experiment; moreover, the reading literature strongly supports the view that faster, rather than slower, readers should be sensitive to lexical phonology. For example, we know that developmentally, a general characteristic of good readers is their superior phonological skills; conversely, dyslexia is almost always associated with phonological weakness (e.g. Goswami & Bryant, 1990; Siegel, Share & Geva, 1995).

However, this finding poses problems for the dual-route approach. It might be argued that slow readers use the assembly procedure to the exclusion of the lexical route, and fast readers make use of additional information from the lexical route. But, for the implemented DRC model, this explanation is the opposite of what actually occurs. Because the sublexical route is slower than the lexical route, it is the slower readers who are more likely to demonstrate lexical effects because there is more time for contributions to emerge from the lexical route. This is precisely the explanation given by Coltheart et al. for Taft and Russell's (1992) slow readers. Speeding up the nonlexical route in order to simulate fast reading would abolish the effect of lexical phonology since an output would be generated before there was time for activation via the lexical route.

A single-route explanation needs to posit two different loci for the reader speed effect. Fast readers have developed a greater sensitivity to more frequently articulated sounds, while slower readers are more dependent on spelling-to-sound mappings within the reading system itself and therefore show a sensitivity to wordlikeness. However, this seems to run counter to the underpinning assumption of the single-route approach which would predict that fast/efficient readers should in theory be more sensitive to orthotactic probabilities than slow/less efficient readers. A two-route connectionist model as proposed by Zorzi et al. (1998) is theoretically promising, because the effects might emerge from fast readers making greater use of their knowledge of the lexical route, but in its current implementation, there is no mechanism for nonwords to activate the lexical route. However, a resonance account of the type proposed by Van Orden and colleagues (e.g. Becker, Goldinger & Stone, 2006; Stone, Vanhoy & Van Orden, 1997; Van Orden & Goldinger, 1994) might offer a more promising approach, in that their covariant learning approach specifically accounts for individual differences. "Different readers… may sample language differently. Each reader has a unique history of covariation among words' spellings, phonology and semantics" (Van Orden & Kloos, 2006, p. 71). From this we might argue that fast readers might well have a highly-integrated system in which effects of lexical phonology might be apparent, whereas slower readers would be more likely to make use of sublexical sources of information, and therefore demonstrate responses affected by N and wordlikeness. In general, it is clear that we need to be alert to the fact that individuals may have different reading styles and abilities (e.g. Paap & Herdman, 1998); individual differences may affect experimental results, but may also be a fruitful avenue for future research.

Wordlikeness

The experiment has given unequivocal evidence that visual wordlikeness affects responses in pseudohomophone naming. Whatever criteria people used to make their ratings affected responses made by other people. The more wordlike items were named more quickly, and in tasks where orthographic factors were likely to be relatively prominent (e.g. in mixed lists and for slower readers) wordlikeness had a particularly strong influence. The findings are not captured in current dual-route theory, and the DRC model actually read the more wordlike items more slowly. But the findings do lend support to the view that readers have extracted information about the spelling-to-sound patterns in their native language, and that they make use of this implicit knowledge when they name letter strings. When presented with wordlike stimuli, readers can switch on the mechanisms that operate during normal reading processes.

When presented with unwordlike stimuli, however, it is unlikely that normal mechanisms come into play. At a coarse-grained level, we knew this already, because rime status has been shown to affect responses. Vanhoy and Van Orden (2001) demonstrated that in lexical decision, pseudohomophones like *jale* produce reliable pseudohomophone effects but items like *stahp* ('stop' in American English) do not. Herdman et al. (1996) demonstrated that pseudohomophones with legal rimes are named faster than those with illegal rimes. But it is clear that rimes are not the only factor in wordlikeness ratings, otherwise pseudohomophones with the same rimes would have been given the same ratings, and they were not (e.g. *dence* = .94, *sence* = .87). It might be argued that base word phonology affects pseudohomophone ratings and accounts for the difference, but with true nonwords we see that *loak* > *moak* (.55, .16), and *cade* > *dade* (.54, .34). Readers are therefore sensitive to factors other than, or additional to, rime. While these factors might be those that have featured in previous research, (bigrams, trigrams, N, length, etc) there might also be as yet undefined orthotactic and graphophonemic relationships that are captured in the 'wordlikeness' measure.

Summary

Overall, readers do not show pseudohomophone and frequency effects; wordlike nonwords and pseudohomophones can be read in the same way. However, faster readers do show effects of lexical phonology because they read pseudohomophones faster than nonwords, and show an effect of base word frequency. By contrast, weaker readers do not show frequency effects, but they do manifest sensitivity to the orthographic variables of N and wordlikeness. The best account of these findings is a system that incorporates both lexical and sublexical knowledge and that learns spelling-to-sound patterns from exposure to a body of print. The lexical route, or process, is as efficient as, if not faster than, the nonlexical route and therefore the current formulation of the DRC model does not capture the human data. In addition, the rules embodied in the DRC GPC route do not appear to be those implicitly used by human readers, and therefore the DRC cannot capture the effects of orthographic wordlikeness. A dual-route account allowing for an efficient set of spelling-to-sound mappings and a fast nonlexical route can account for these findings; such an account might be that postulated by Zorzi et al., and the resonance account proposed by Van Orden and colleagues also seems to offer potential mechanisms for explaining these results.

2.4 Experiment 2:
What does cAsE aLtErNaTiOn do to pseudohomophone naming?

Evidence from Experiment 1 suggests that pseudohomophone and base word frequency effects only occur for fast readers; and these are not post-assembly phenomena, but are best explained in terms of a contribution from lexical representations. It is likely that the use of more wordlike items, with legal orthographies, is responsible for the clear findings of Experiment 1, and supports the position that results of earlier work are confounded because of unwordlike stimuli. While the effects of frequency and N come and go as a function of list order and composition, the influence of wordlikeness is much more stable. If wordlikeness is a measure of implicit orthotactic knowledge, including the relative co-occurrence of individual letters and letter-groupings, we can predict that disrupting the visual appearance of the stimuli will affect processing so that it decreases in importance as a predictor of naming latencies. Such a disruption can be achieved through cAsE aLtErNaTiOn.

Case mixing, in various forms, has long been known to disrupt word recognition, relative to when participants read words in single (upper or lower) case (e.g. Mason, 1978, Smith, 1969). For nonwords, support for the view that the wordlikeness effect seen in Experiment 1 will disappear with case mixing comes from an experiment reported by Campbell and Mewhort (1980). They showed that although recall performance was facilitated for letter strings that approximated to the orthographic structure of English, the effect was reduced by increasing the spaces between letters. They argued that letter spacing had the effect of damaging the mechanism that parses individual letters into graphemic units. A similar view was taken by Pring (1981) who found that disrupting the graphemes she named 'functional spelling units' had a greater effect than if the units remained intact. Thus, ChuRCH generated larger processing costs than CHurCH, and pseudohomophones such as CHerCH activated base word phonology in lexical decision while cheRCH did not.

While it is clear that case mixing disrupts normal orthographic processing, the underlying reason for its damaging effect is not clear. Responses are slowed overall, presumably because the number of possible patterns from which each letter must be recognised is doubled as opposed to pure case presentation (e.g. Paap, Newsome, & Noel, 1984). But a general slowing is not the only effect; case alternation also changes the normal processes for converting print to sound. Responses are also more error prone for both words and nonwords, (e.g. Allen, Wallace, & Weber, 1995; Besner & Johnston, 1989) and this is likely to be because it disrupts the automatic parsing process that converts single letters to graphemic units. There are various potential sources of this disruption: it may occur because letters are masked (e.g. lateral masking from uppercase to lower case; Besner & Johnston, 1989), or because it introduces inappropriate grouping between letters with the same size and case (e.g. ArEa provides a conflicting grouping of A and E); so-called ‘transletter features’ such as the distortion of spaces between letters may also be responsible (Mayall, Humphreys & Olson, 1997). Reviewing the evidence, Mayall and Humphreys (1996) suggested that case mixing might either affect a relatively early stage in letter encoding before the construction of the code used to access the lexicon (i.e. an abstract letter code), or alternatively it might disrupt the nonlexical route. They excluded a third possibility, the disruption of word shape, because high frequency words, with very familiar shapes, are not particularly disrupted with case mixing.

Aim and predictions

If case mixing effects in naming disrupt the processes that occur when familiar letter combinations are processed, and if wordlikeness indexes reader sensitivity to the relative frequency of these graphemic patterns, then it is likely that wordlikeness effects will be reduced for case mixed stimuli. Further, if disrupting visual appearance by case alternation has the effect of reducing responses based on phonology (cf. Pring, 1981), then we should also expect that the pseudohomophone effect shown by fast readers will no longer be apparent. However, if the speeded pseudohomophone effect seen in the pure list presentation in Experiment 1 arose from response criterion-setting by participants, then there is no reason why a similar effect should not occur again under case-alternated conditions.

Finally, case alternation is a useful way to test the single-route theorists' view that pseudohomophone and frequency effects (insofar as they exist in naming) occur at a post assembly stage. If this is true, then pseudohomophone and frequency effects should still be detectable for fast readers as in Experiment 1. If case alternation disrupts these effects, then it is likely that they are not occurring at articulation, but earlier on in the process of assembling a phonological representation.

2.4.1 Method

The procedure and stimuli were identical to those in Experiment 1. 40 University of Bristol students who had not taken part in the first experiment were recruited. The only difference from Experiment 1 was that the pseudohomophones and true nonwords were presented in case alternated form. Nonwords were created in both possible forms of case alternation e.g. BeEm and bEeM, except for a few instances where this would have created an ambiguous form (e.g. ClEeR), in which case, only the unambiguous presentation was used (e.g. cLeEr). Participants were presented with an equal number of lower case/upper case initial consonants.

2.4.2 Results

Overall mean latencies and standard deviations for correct responses to the critical items are shown in Table 6; results from Experiment 1 are also included. Analyses were performed on trimmed data; scores more than 3 standard deviations below or above an item mean were classed as outliers and excluded.

Table 6
Mean reaction times, standard deviations and errors for pseudohomophone naming in pure and mixed lists.

	Pseudohomophones			Nonwords		
	Mean RT (ms)	S.d.	Errors (%)	Mean RT (ms)	S.d.	Errors (%)
Pure lists overall	586	93	5.2	580	88	8.5
Expt 1	*521*	*39*	*7.4*	*539*	*45*	*8.1*
Difference Ex 1 and 2	*65*	*54*	*-2.2*	*41*	*43*	*0.4*
Mixed lists overall	549	103	5.7	551	105	6.9
Expt 1	*514*	*41*	*4.0*	*518*	*40*	*6.9*
Difference Ex 1 and 2	*35*	*62*	*1.7*	*33*	*65*	*0*

Case alternation had the effect of increasing naming latencies (Figure 8) and the manipulation selectively damaged pseudohomophones over nonword processing in pure but not mixed lists (pure lists: $F2\ (1,102) = 14.6$, $p < .001$; for mixed lists, n.s.). There was clearly little difference in reaction times between pseudohomophones and nonwords in either mixed or pure lists; analyses were all non significant. The analysis of interest is therefore to determine whether the speeded pseudohomophone effect seen in Experiment 1 was also detectable here.

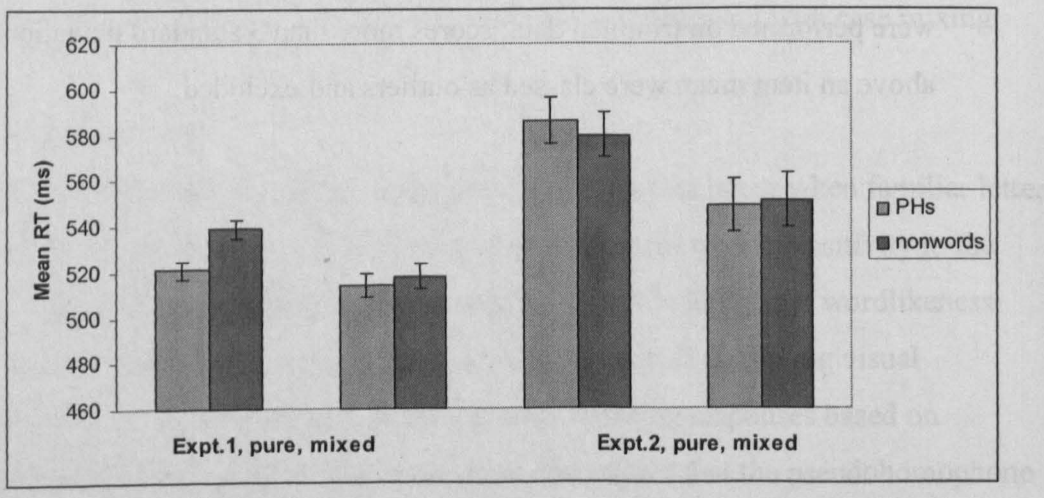


Figure 8. Pseudohomophone and nonword naming times, and standard error of the mean, by list type, for Experiments 1 and 2.

Table 7

Mean reaction times, standard deviations and error rates to pseudohomophones and nonwords as a function of list presentation order.

	Pseudohomophones			Nonwords			*RT Diff*
	Mean RT (ms)	S.d.	Errors (%)	Mean RT (ms)	S.d.	Errors (%)	
PHs 1st	622	85	6	593	86	7	*29*
PHs 2nd	550	86	6	568	91	10	*-18*

Table 7 shows that pseudohomophones were named 18 ms faster than nonwords when read second, a difference that was significant by subjects and marginal by items, (t1 (9) = 2.5, p = .036; t2 (102) = 1.8, p = .074), a finding similar to that in Experiment 1. In addition, pseudohomophones were named 29 ms more slowly than nonwords when read first, a naming disadvantage not found in Experiment 1 (t1 (9) = 2.4, p = .043, t2 (102) = 3.4, p = .001). See also Figure 9.

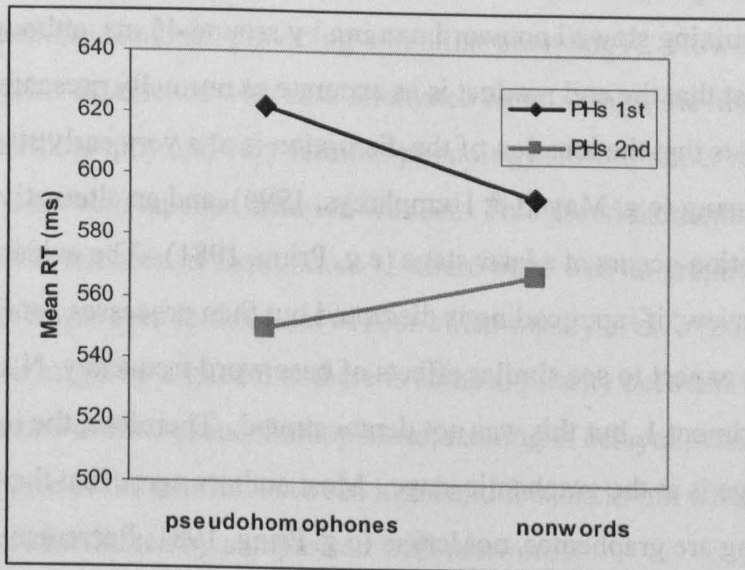


Figure 9. Pseudohomophone latencies by presentation order (pseudohomophones first or second).

Wordlikeness, base word frequency and N

Although there were significant pseudohomophone effects in pure lists, reaction times did not correlate with base word frequency in either list presentation condition.

There was no correlation with N in any type of list. Wordlikeness, however, correlated with latencies, in mixed lists only ($r(52) = -.31$, $p = .03$).

Reader speed

There was no pseudohomophone effect for the 10 fastest readers. For the fastest readers in mixed lists, latencies showed a marginal correlation with wordlikeness ($r = -.263$, $p = .06$). RTs did not correlate with base word frequency for fast readers in any list condition. Slow readers showed no correlation with base word frequency, N, or wordlikeness in any type of list presentation.

Summary

Reaction times were slowed overall, as typically occurs in case alternated experiments. List composition and presentation order affected responses, as with Experiment 1, although this time pure list naming times were longer overall. There was no reader speed effect. Wordlikeness correlated with latencies in mixed lists.

2.4.3 Discussion

Case mixing slowed nonword naming by around 45 ms, although the error data suggest that the end product is as accurate as normally presented items. One view suggests that the location of the disruption is at a very early stage of input coding processing (e.g. Mayall & Humphreys, 1996) and an alternative view suggests that disruption occurs at a later stage (e.g. Pring, 1981). The evidence here supports the latter view; if input coding is disrupted but then processes continue normally, one might expect to see similar effects of base word frequency, N and wordlikeness as in Experiment 1, but this was not demonstrated. Therefore, the indication is that damage is at the graphemic stage. Most authors agree that the functional units of reading are graphemes, not letters (e.g. Pring, 1981; Peereman & Content, 1997; Rey, Ziegler & Jacobs, 2000; but see Rastle & Coltheart, 1998, for an alternative view), and data from both skilled readers and acquired dyslexics (Caramazza & Miceli, 1990; Derouesné & Beauvois, 1979; Rey et al., 2000) have been taken to indicate the existence of a dissociable graphemic parsing process that extracts orthographic units for subsequent lexical processing. Case alternation disrupts this automatic parsing process and forces readers of all types to rely more heavily than usual on letter-level analysis and/or serial processing of letters (Mayall et al., 1997;

Mayall, Humphreys, Mechelli, Olson & Price, 2001). Fast readers were particularly disadvantaged by case alternation; Experiment 1 revealed a pseudohomophone and base word frequency effect for these readers, suggesting an efficient and highly-integrated system. However, when fast readers were required to process letter strings without the normal routines that allowed graphemes to be quickly activated, lexical activation was compromised.

As with Experiment 1, there was no difference between pseudohomophone and nonword naming times in mixed lists, but there were pure list order effects. As before, pseudohomophones were named faster after a list of nonwords, but, in addition, were also read more slowly when presented before the nonword list. As before, there was no concomitant base word frequency effect, so it is unlikely that lexical representations were being accessed in these difficult naming conditions. A response-setting criterion is again a more appropriate explanation. When participants were faced with a list containing case alternated pseudohomophones, the naming output was of a familiar phonology. When participants had already read a list containing nonwords, this familiarity had the effect of speeding responses (perhaps by virtue of auditory feedback providing a familiar phonology). However, when there was no prior experience with case alternated letter strings, the mismatch with very unusual orthography and very familiar phonology slowed rather than speeded responses, so a slower response criterion was set. This pseudohomophone disadvantage was not seen in Experiment 1, where there was no graphemic disruption; but it was seen in three out of four of Borowsky et al.'s (2002) experiments. It might be argued that there is some similarity between wordlike/case-alternated and unwordlike pseudohomophones; naming is delayed when they are read first, because of the conflict between input and feedback; but when they are read second, they are facilitated by comparison with nonwords.

Evidence for the challenging and unusual nature of the task is shown in that individual differences in readers were no longer detectable. Case alternation effectively forced participants to adopt alternative processing strategies; the large standard deviations suggest that they may be considerable variability in these strategies. The base word frequency effect for fast readers disappeared, suggesting that orthographic disruption meant that their normal default strategies leading to lexical phonological activation could no longer be employed. This is perhaps not surprising given that an inability to form normal grapheme-to-phoneme associations is linked with impaired phonological activation and is a feature of dyslexia (e.g. Siegel et al., 1995). For all readers, however, wordlikeness was still present in mixed lists as an influencial variable. This indicates that even in this disrupted condition, readers made use of their orthographic knowledge at some level – this might be sensitivity to individual, or initial, letters. It was thought that case alternation would abolish the wordlikeness correlation in particular, as it was thought to be essentially a measure of visual appearance, so it is interesting that where all other effects have been obliterated there is still some remaining aspect of the pseudohomophones that, even when presented in mixed case, makes them more or less like real words.

Finally, the evidence is not encouraging for the single-route explanation favoured by Seidenberg et al. Their view was that, if pseudohomophone and frequency effects exist, they are the result of post-assembly activity at the point of articulation. On the basis of the evidence above, pseudohomophone and frequency effects should still be apparent, because the articulatory units are not affected by the orthographic input; i.e. the articulatory units for /breın/ have no knowledge of whether the input orthography was upper case, lower case, or a mixture.

2. 5 Summary and conclusion

Experiment 1 showed that readers are sensitive to wordlikeness and that individual differences and list presentation factors affect naming responses. There was no clear demonstration of an overall effect of facilitation for pseudohomophones, and it seems likely that when participants are presented with *soke* and *sote*, they deal with them in the same way; although this may not be the case for fast readers. Case alternation abolished the various patterns of effects found in Experiment 1, although there was a residual effect of wordlikeness, indicating that even under difficult reading conditions readers were able to make use of some of their knowledge of spelling-to-sound regularities. The disruption prevented graphemes from being processed in the normal way, and this prevented skilled readers from accessing lexical phonology. If researchers use nonwords that do not make use of familiar graphemes, then it is arguable that they are effectively forcing participants to read like slower or possibly weaker readers, and it should come as no surprise when phonological effects are hard to find. Where they have occurred, phonological effects may be the result of experimenter instructions, list or stimulus construction, and/or readers adopting problem-solving strategies to deal with atypical letter sequences.

Chapter 3

The effects of visual wordlikeness in lexical decision experiments

A Czech went to the optician, who showed him an eye chart and asked, "Can you read the bottom line of letters?" "Read it?" said the Czech; "That's my brother!"

In the previous chapter, it was shown that when wordlike stimuli are used, there is no overall difference between pseudohomophone and nonword naming latencies. The apparently speeded naming when pure lists were read second was a strategic effect, and the co-occurring pseudohomophone and base word frequency effect seen with the 15 fastest readers was an indication of reading processes specific to this group of readers. For the majority of readers, however, pseudohomophone and base word frequency effects do not occur in naming. It was argued that the best model to account for these findings needs to include two separate routes or processes, one assembled and sublexical and one lexically-based. Individual differences in reader skill lead to a greater or lesser integration of these two routes. In order to explain the pseudohomophone and base word frequency effects, feedback connections as well as feedforward, must be implicated, and there is greater connectivity for the more skilled readers.

In addition, it was argued that the findings reported in the previous chapter are more convincing than those of previous experiments where the use of unwordlike nonwords means that it is far more difficult to interpret how participants might have been processing the stimuli. In support of this view, results from Experiments 1 and 2 showed that participants were sensitive to the perceived wordlikeness of nonwords. This seemed to be primarily an orthographic variable in that the slower participants, who were sensitive to N, were also sensitive to wordlikeness. Latencies of the faster participants did not correlate with these variables; again adding support to the idea that slower participants were reliant on an assembled sublexical route, while faster participants gained additional support from a lexical route. Wordlikeness is clearly a robust and important variable, because it was the only one that survived the case mixing procedure used in Experiment 2, while pseudohomophone, frequency and N effects all disappeared.

3.1 Lexical decision

An alternative approach to investigating the role of phonology in visual word recognition is to use the lexical decision task, which is a useful paradigm because it does not require the additional articulatory processes involved in naming experiments so is arguably a more appropriate task when seeking to uncover phonological effects. The two processes are comparable since lexical decision items are very similar to times in normal reading, so it is legitimate to generalise from these tasks to word recognition process in reading (Schilling, Rayner & Chumbley, 1998). In visual lexical decision, participants are asked to decide whether a letter string looks like a word or not, or whether it spells a word. Phonological lexical decision tasks, on the other hand, require participants to decide whether an item sounds like a word. Phonological effects are thus clearly predicted in this task, but in visual lexical decision, judgements can be made on the basis of orthography alone, so any phonological effects can be taken as evidence of automatic phonological activation of lexical entries and thus support the 'strong' phonological model of visual word recognition. As Frost (1998) puts it: "one can refute the weak phonological position by demonstrating that phonological recoding is present even in tasks in which it is not required or hinders performance." (p. 76).

In visual lexical decision the standard finding is that, when presented with words, pseudohomophones and nonwords, participants are faster to accept words than to reject all types of nonword, and take longer to reject pseudohomophones than control nonwords. Additionally, error rates are higher for pseudohomophones than for control nonwords (e.g. Coltheart et al., 1977; McCann & Besner, 1988; Rubenstein, Lewis & Rubenstein, 1971; Seidenberg et al., 1996; Ziegler et al., 2001). These findings would appear to offer convincing support for the 'strong phonology' school, but they are also straightforwardly explained by models based on the 'weak' assumption. For example, in the dual-route model, the grapheme-to-phoneme route generates phonological activity that feeds back to whole word representations, and this lexical match delays processing. Thus, 'no' responses to *brane* are slower than 'no' responses to *brate*, because the lexical status of the former signals an incorrect 'yes' response; *brate* does not generate such conflict and is therefore faster. An alternative explanation is offered by single-route theorists such as Seidenberg, who suggest that the fact that a pseudohomophone's phonological code is that of a word interferes with making a nonword response because it activates semantic information associated with the word. A compromise position between weak and strong theories has recently been advanced by Rastle and Brysbaert (2006), who suggest that the recognition of printed words in lexical decision is based largely on an analysis of phonological representations, the activation of which is constrained by orthographic information. On this account, activation in the phonological lexicon is stronger for words, by virtue of summed information from both orthography and phonology, than for pseudohomophones, which have no support from the orthographic lexicon.

If pseudohomophones activate phonological lexical representations, one might expect to see base word frequency effects, but they are generally not found in visual lexical decision. Null effects were reported by McCann et al. (1988), and Seidenberg et al. (1996), although Ziegler et al. (2001), reported an effect on latencies in German, and Van Orden et al. (1992), identified frequency effects in errors. On the basis of their findings, McCann et al. argued that the phonological lexicon is not frequency-sensitive. An alternative suggestion is that a base word frequency effect does not occur because the task can be accomplished by readers carrying out a familiarity check based on an analysis of the letter string's spelling. In the case where unwordlike items are presented, this is arguably a plausible and effective strategy; unfamiliar letter combinations such as *ehj* and *terhn* can be swiftly discounted purely on the basis of their bizarre orthographic patterns. Using wordlike letter strings enables a more valid investigation of the processes involved in normal reading, especially the status of lexical phonology and the interesting possibility that phonological representations, here conceptualised for simplicity's sake in the notion of a 'phonological lexicon', are not frequency sensitive.

Computer modelling of visual lexical decision

Theoretical explanations of performance in visual lexical decision can be strengthened if a computer model demonstrates similar performance to humans and such data have been reported by Coltheart et al. (2001), and Jacobs et al. (1998). Although the DRC and MROM-P models are slightly different, as far as this task is concerned, the two models make similar predictions based on similar assumptions; words elicit responses by means of an activation criterion applied to individual lexical entries in an orthographic lexicon; 'no' responses occur when a deadline passes. This deadline is variable, and moves back when the global activation in the lexical system is high. Words are therefore faster than nonwords, and high frequency words are facilitated. The pseudohomophone effect is explained as feedback activation from the assembly route to the orthographic lexicon via the phonological lexicon. Occasionally, a pseudohomophone may drive the activation of an individual word unit above the critical activation threshold and if this happens, the model will give an incorrect 'yes' response; true nonwords are unlikely to do this.

According to Ziegler et al., (2001) however, both the DRC and the MROM-P make incorrect predictions about what actually happens in lexical decision. Because pseudohomophones based on high frequency words generate more activity in the phonological lexicon, this generates greater conflict between the two routes, and therefore pseudohomophones derived from high frequency base words take longer to reject. However, Ziegler et al., Experiment 1 (in German), and Van Orden et al., (1991) found faster decision latencies for pseudohomophones derived from higher frequency base words. In order to account for this finding, Ziegler et al. suggested that an additional verification mechanism needed to be incorporated into this type of model. Orthographic representations for high frequency words would be more stable than for low frequency words, which would mean that verifying the spelling of pseudohomophones based on high frequency words would be facilitated over low frequency items.

A system that integrates activation and verification was proposed by Van Orden and Goldinger (1994) and Van Orden, Jansen op de Haar and Bosman, (1997) in the resonance/coherence model. In this account, a visually presented stimulus word or nonword activates visual units, phonological units, and semantic units, but this initial activation is not sufficient to distinguish words from nonwords. Lexical decisions are based on the coherence of activation, which is a measure of self-consistent feedback, matching bottom-up activation across different families of units. Nonwords produce bigger mismatches than pseudohomophones because the latter find additional matches in connections between phonology and semantics, so nonwords are rejected more quickly. This account can also explain the frequency effect. As top-down expectations are matched, or verified, against incoming bottom-up activation, pseudohomophones create a mismatch because of differences in spelling. Because frequent base words provide stronger top-down expectations than infrequent base words this mismatch is amplified for pseudohomophones derived from high frequency base words; the stronger the mismatch, the faster the 'no' responses.

Individual differences in reader speed/skill

The previous chapter explored the idea that individual differences in reading may have contributed to the inconclusive results in naming experiments, and it is also possible that such difference might have affected results in lexical decision experiments. In Experiment 1 a difference between fast and slow readers was noted, such that fast readers showed co-occurring pseudohomophone and frequency effects. The performance of slow readers, on the other hand, was more influenced by orthographic variables. It was argued that fast readers had well integrated sublexical and lexical routines, while slow readers were making more use of sublexical processes. How might individual differences manifest themselves in lexical decision? In the visual task, finding that a nonword is a pseudohomophone hampers performance, so one might expect skilled readers to selectively disregard phonological information and to concentrate on the orthography of the nonword items. This enables all nonwords to be judged in the same way, irrespective of their homophonic status. Slow readers, on the other hand, might be less able to identify and selectively make use of effective strategies and they might therefore be less able to disregard the conflicting information from a pseudohomophone's base word. A pseudohomophone effect is more likely to occur with slow readers, therefore. This pattern of differences was detected in visual lexical decision by Seidenberg et al. (1996) who found that the fast participants showed no pseudohomophone effect, but the slow and medium readers took longer to make decisions to pseudohomophones. They also suggested that fast readers can ignore phonological information, while the slow readers are hampered by it. However, the data are also capable of being interpreted from a dual-route perspective: for example, the bypass hypothesis assumes that skilled readers can bypass phonological information through the use of direct orthographic access; beginning or less skilled readers rely exclusively on phonological processes (Baron, 1973; Doctor & Coltheart, 1980). An alternative explanation in terms of activation – verification as outlined above would be that faster readers have more efficient verification procedures and more accurate orthographic representations.

Phonological Lexical Decision

More robust frequency effects co-occurring with pseudohomophone effects are reported in the alternative form of lexical decision tasks, when participants are asked to decide whether a nonword *sounds* like a real word. In phonological lexical decision tasks, Grainger et al., (2000), McCann, Besner and Davelaar (1988), and Taft and Russell (1992) reported that participants were faster to say 'yes' to pseudohomophones than they were to say 'no' to nonwords, and that items based on higher frequency words generated faster decision times and fewer errors than those based on lower frequency words. The usual explanation is that, since pseudohomophones activate real word matches, they are responded to more quickly than nonwords, which do not activate matches, and base word frequency effects occur as a result of activation in a frequency-sensitive phonological lexicon – more frequent items elicit a match more quickly than less frequent. McCann et al., however, argued for a post-lexical locus for the effect, with participants attending to the output of a familiarity discrimination mechanism, operating on the phonological code assembled from the letter string. This mechanism is redundant in naming and visual lexical decision, so no base word frequency effects are seen in those tasks. But these data should also be treated with caution; Grainger et al.'s stimuli were only one letter different from their French, disyllabic, base words, so it is possible that the source of the frequency effect stemmed from orthographic activation rather than phonological; and Taft and Russell's findings were non-significant by items. McCann et al. reported a significant negative correlation between latencies and base word frequency but inspection of the scatterplot suggests that four or five low frequency items generated exceptionally long latencies; removal of these might well have reduced the strength of the correlation. Individual item data were not reported so it is not possible to confirm this.

Summary

Findings in lexical decision tasks are slightly more coherent than those obtained from naming experiments. However, although pseudohomophones are clearly responded to differently from words and nonwords, establishing a base word frequency effect has been more difficult and this is problematic for models that invoke a frequency-sensitive phonological lexicon. As was argued with the naming experiments, results that are difficult to interpret may have been confounded by the orthography of the nonword stimuli. Therefore, the next two experiments aimed to replicate standard visual and phonological lexical decision tasks with more wordlike items, previously rated for wordlikeness.

3.2 Experiment 3:

Visual lexical decision with wordlike stimuli

Aim and predictions

The default prediction in visual lexical decision is for words to be responded to more quickly than all types of nonword, and for pseudohomophones to elicit slower 'no' decisions than true nonwords. Since Experiments 1 and 2 demonstrated that the phonological lexical status of wordlike pseudohomophones does not necessarily affect naming times, it is possible that there will be no difference between pseudohomophone and nonword stimuli in visual lexical decision. Since the new, more wordlike, pseudohomophones did not elicit a frequency effect in the naming experiments, there was no reason to suppose that frequency effects would be apparent in this task. However, since participants in the naming experiments had shown sensitivity to wordlikeness, it was likely that responses to more wordlike items would be faster than to less wordlike items.

3.2.1 Method

Participants

30 students at the University of Bristol who had not taken part in the previous experiments participated in return for partial course credit.

Apparatus

The display appearance and timing was the same as for Experiments 1 and 2, except that in this case the target letter string remained visible until the participant responded. The equipment set-up was similar to that used in Experiments 1 and 2, but with a keypress rather than a spoken response.

Stimuli and design

The 52 pseudohomophones used in Experiments 1 and 2 were used as the target stimuli, with the replacement of three items that most participants had not recognised as pseudohomophonic. The 52 comparison nonwords were also used, together with 104 monosyllabic words of 4 – 6 letters chosen to have the same frequency range as the pseudohomophones (see Appendix B). Stimuli were randomly presented with the constraint that the same response should not be given more than three times in a row.

Procedure

Participants were informed orally and in writing that they would see letter strings presented on the computer screen in front of them, and that their task was to decide whether each letter string was an English word or not. Responses were given by a keypress to one of two buttons on a customised response box. A practice list of 20 items was given to check that the participants understood the instructions. The experimental session was broken into four blocks of 52 trials, with rest pauses between blocks. Within these blocks, trials were run in a continuous sequence, with a fixation cross displayed for 500 ms followed by a blank screen for 150 ms and the stimulus target presented until a response was given. There was an interval of 1.5 sec between the response to one target and display of the fixation cross for the next. Total testing time was around 20 minutes.

3.2.2 Results

Trials were excluded from the analysis if an incorrect response was made, or if the reaction time for the trial was more than 2.5 standard deviations from the participant's mean reaction time for that condition. This procedure led to three nonword items scoring as errors on 57% of the responses, and means for these three items were excluded from the reaction time analysis (pseudohomophone *churp* and nonword controls *raim* and *storn*). The analyses that follow were carried out on the trimmed data, which showed normal distributions in all three conditions.

Table 8
Decision times in ms, standard deviations, and errors in visual lexical decision.

	RT mean (ms)	S.d.	Errors (%)
Pseudohomophones	715	61	1.4
Control nonwords	688	65	2.1
Words	583	68	1.6

Table 8 indicates that the normal pattern of results in lexical decision occurred, in that response latencies to words were fastest, and that pseudohomophone rejections were slower than those to nonwords. Analyses of variance by subjects and by items were conducted to investigate these differences. Planned comparisons using *t* tests were then used to investigate differences between pseudohomophone and nonword reaction times. The repeated measures analysis of variance by subjects showed that there was a significant difference between means: $F1\ (2, 58) = 76.23$, $p < .001$. The one-way analysis of variance by items showed that there was a significant difference between the means: $F2\ (2, 202) = 86.37$, $p < .001$. The paired samples *t* test showed that participants rejected control nonwords significantly more quickly than pseudohomophones: $t1\ (29) = 2.98$, $p = .006$. An independent samples t test was used for the by items analysis, and this was also significant: $t2\ (99) = 2.1$, $p = .04$.

The above analysis was repeated, using error scores as the dependent variable. The ANOVA was significant by subjects: $F(2, 58) = 11.86$, $p < .001$, and approached significance by items: $F(2,202) = 2.5$, $p = .08$. Planned comparisons using t tests showed that the differences between pseudohomophones and nonwords were significant: $t1\ (29) = 6.95$, $p < .001$; $t2\ (99) = 2.42$, $p = .02$.

Reaction time and errors were correlated with N, base word $\log_{10}$ frequency and wordlikeness. Latencies did not correlate with frequency, but the positive correlation with wordlikeness indicated that the more wordlike items generated longer decision times (see Table 9). Errors correlated with frequency and wordlikeness; the relationship with frequency was negative, while the wordlikeness correlation was positive.

Table 9
Correlations between pseudohomophone response latencies with base word frequency, N and wordlikeness.

	Base word frequency	N	Wordlikeness
Reaction time (ms)	r = -.09	r = .18	r = .378**
Errors	r = -.30*	r = -.09	r = .379**

** = p < .05, ** = p < .001*

Participant speed

Participants were divided into fast, medium and slow on the basis of their nonword (pseudohomophone and true nonword) response times (see Figure 10). There were significant main effects of group ($F(2, 297) = 178.51$, $p < .001$) and stimulus type ($F(1, 297) = 7.14$, $p = .008$). The interaction was not significant ($F(2, 297) = 2.52$, $p < .1$). The effect was 4 ms for fast participants, 20 ms for medium participants and 58 ms for slow participants; the difference between pseudohomophones and nonwords was only significant for the slow participants ($t1\ (9) = 3.8$, $p = .009$; $t2\ (99) = 2.32$, $p = .022$).

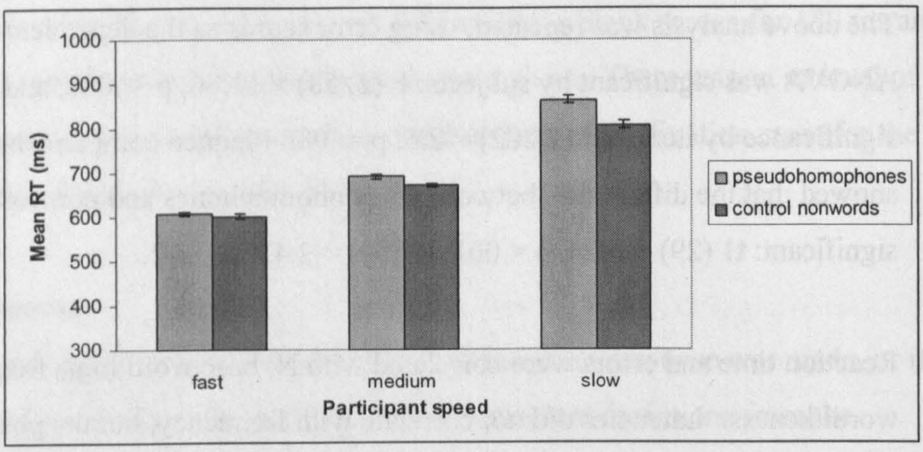


Figure 10. Mean lexical decision latencies and standard errors of the mean as a function of participant speed.

Table 10

Percentage errors by participants

	Pseudohomophones	Nonwords
Fast	3.7	5.5
Medium	4.8	7.0
Slow	5.6	8.3

Error rates showed a systematic pattern across fast, medium and slow participants for both types of nonword. Fast, medium and slow participants' latencies and errors to pseudohomophones were correlated with N, frequency and wordlikeness. Only wordlikeness correlated with reaction time such that more wordlike items were responded to significantly more slowly for all groups. Wordlikeness and frequency correlated with errors for the medium and slow groups. N was not significant for reaction time or errors.

Table 11

Correlations between pseudohomophone responses with frequency and wordlikeness by participant speed

		Wordlikeness	Base word frequency
		r value	r value
Reaction time (ms)	Fast	.54, **	-.04
	Medium	.31*	-.05
	Slow	.29*	-.18
Errors	Fast	.17	-.18
	Medium	.27, p < .1	-.24, p <.1
	Slow	.45**	-.29*

** = p < .05, ** = p < .001*

3.2.3 Discussion

Experiment 3 replicated previous findings in visual lexical decision, in that words were responded to more quickly than all types of nonwords, but pseudohomophones elicited slower 'no' responses than control nonwords. If pseudohomophones generate a phonological representation that is that of a real word, this will cause conflict and therefore slow response time. Such conflict is also likely to produce more errors than to control nonwords, and this pattern was apparent in the untrimmed data (pseudohomophones, 1.3%; nonwords, 0.5%), which is the appropriate place to look for the effect, since the final trimmed dataset included time-outs rather than true errors. Latencies for the pseudohomophones did not correlate with base word frequency, replicating the findings of McCann et al. (1988), and Seidenberg et al. (1996) but failing to support findings reported by Ziegler et al. (2001). If the phonological lexicon is frequency sensitive, it would be expected that decisions to pseudohomophones would show a frequency effect, such that items based on high frequency words would take longer to respond to since these items would generate more activation and more conflict with the input. However, Ziegler et al. found the opposite, in that items based on high frequency base words were faster to respond to. Taken together, their findings plus the null effects from McCann et al., and Seidenberg et al., together with those from the current study, do not offer support for the notion of a frequency sensitive phonological lexicon, as postulated by models such as the DRC and MROM-P.

It might be argued that a familiarity check based on the spelling of the word, as suggested by McCann et al., would be sufficient for the decision-making processes in this task, and this would not generate frequency effects; but in this case it is difficult to see why there should be a difference between pseudohomophones and nonwords since both types of stimuli could be checked for familiarity, and if orthography were the only deciding factor, then there would be no difference in decision times. Therefore lexical phonology must have some part to play in visual lexical decision, and its role can perhaps be uncovered by an investigation of the error analysis. This study has shown that stimuli based on low frequency words were more likely to cause errors, thus replicating the findings of Van Orden et al., 1992, and Ziegler et al., 2001. (It is worth noting that McCann et al. and Seidenberg et al. also obtained negative, although non-significant, correlations with base word frequency in their error analyses.) If the lexicon were frequency sensitive it would be expected that items based on high frequency base words would cause more activation, and therefore generate more errors, than items derived from low frequency base words. So the error data also indicate that the phonological lexicon is not sensitive to frequency.

So where might we look for the locus of the frequency effect? A plausible mechanism might be that phonology activates the orthographic representation of the base word, and the frequency effect is to be found in feedback from the orthographic lexicon. High frequency orthographies have stronger mappings from sound to spelling than low frequency, so are more likely to be responded to correctly; low frequency words are less strongly mapped and so are more likely to generate errors. See Figure 11 – the connections between the orthography and phonology of the higher frequency word *brain* are strong, whereas the links between a lower frequency word like *crane* are weak. It follows that *brane* is less likely to elicit a false positive than *crain*. This is essentially the explanation that would be offered by the resonance/coherence model; a pseudohomophone activates connections between spelling, phonology, and also semantic units, but creates a mismatch because the bottom-up orthography does not match the top-down phonology. Frequent base words provide stronger mismatches than infrequent base words, so the mismatch is amplified for pseudohomophones derived from high frequency base words. Errors, therefore, are less likely to high frequency items, because of the amplified mismatch, and they are more likely to occur for low frequency items, because the mismatch is not so strong. We have seen that *churp* was wrongly accepted as a real word by more than half of the participants; it is likely that students are simply unsure of the spelling of this low-frequency word.

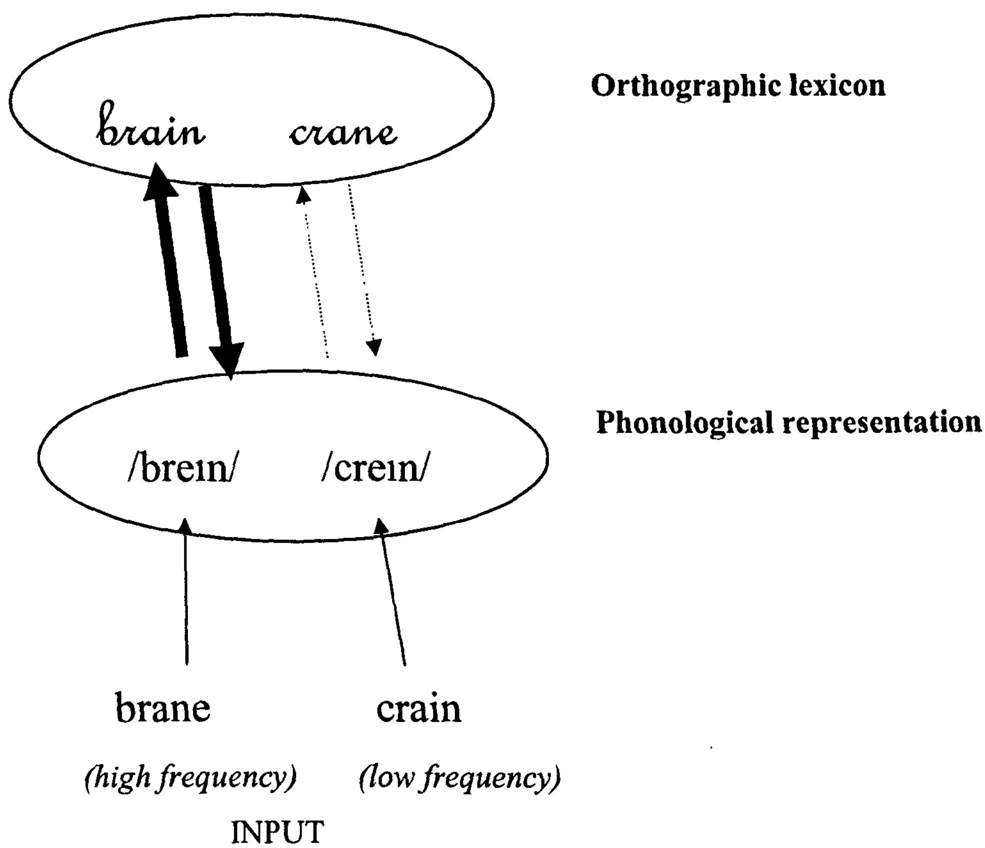


Figure 11. Strong orthography-phonology links for higher-frequency words mean that a pseudohomophone input is less likely to generate a false positive; links for low frequency items are more insecure.

To explain the frequency effect in errors but not in latencies, we would have to assume that most of the time, correct rejections are carried out on the basis of an unfamiliar orthography coupled with activation in a frequency-neutral phonological lexicon. Phonological activation is sufficient to delay responses; but frequency effects only emerge when the base word orthography is also activated. Different experimental conditions and stimulus sets may activate orthography to different degrees - McCann and Besner's items and Seidenberg et al.'s contained many unwordlike items, so are arguably less likely to have activated base word orthography than Ziegler et al.'s and Van Orden's pseudohomophones, which were all just one letter different from the base word. This argument is supported by the observation noted in the Results section, that nonwords *raim* and *storn* both generated high error rates; presumably both these nonwords activated real words ***rain*** and *storm*.

Reader skill/ speed

The findings can be further interpreted in the light of individual differences in reading speed. Overall, these replicated the results reported by Seidenberg et al., (1996), in that the fastest participants did not show a pseudohomophone effect. This supports the view that fast participants can selectively bypass or suppress redundant phonological information, in order to concentrate on orthography. Fast readers also made more accurate decisions, which would be expected if they were able to suppress or ignore the spurious activation of phonology that increased latencies and errors for the slower readers. It was seen that all readers demonstrated an effect of wordlikeness, such that more wordlike items took longer to respond to, but this was strongest for the fastest readers. Since wordlikeness is assumed to be primarily an orthographic variable, this lends support to the view that fast readers relied more on orthographic information to make a decision; and when they encountered a letter string that could very plausibly be a real word, it slowed them down more than for the other groups. However, fast readers did not make errors based on more wordlike items, while slower readers did. If fast readers were carrying out a spelling verification, and if their orthographic lexicon contains very stable spelling representations, then it would be expected that they would take longer to check wordlike items, but they would not make errors to them. Slow readers on the other hand might be assumed to have less stable spelling representations, and would therefore be more likely to make errors to items that looked plausible, particularly if they were unable to suppress additional information coming from the phonological lexicon signalling an incorrect 'yes' decision.

Wordlikeness

A wordlikeness effect was predicted, since Experiments 1 and 2 had shown clear effects in naming. The initial prediction was that responses to more wordlike pseudohomophones were likely to be faster than to less wordlike items, on the basis that more wordlike items had generated faster naming times in the earlier experiments. However, the results showed that, overall, the more wordlike items generated longer decision times, and more false positives. When a pseudohomophone is composed of familiar orthographic components, it becomes more difficult to reject it, because it activates a greater number of possible lexical or sublexical candidates. Unwordlike items do not activate potential matches, so are quicker to reject and generate few errors. With unwordlike stimuli, it is possible that participants base their rejections purely on orthography and if this is the case, we cannot be certain that this represents typical word recognition processing. This point was made by Pexman, Lupker and Jared (2001): if pseudohomophones "are not sufficiently word-like orthographically then participants may be able to make their decisions on the basis of a superficial analysis of the orthography of the stimuli before they have determined whether the stimulus is a word. In that case, the results would tell us little about the processes involved in word recognition" (p. 142). What the results above have shown is that the more wordlike a pseudohomophone, the longer the response time; the limited information offered by the less wordlike items enables quicker and more accurate rejections. Plausible as this account seems, it should be noted that some research has used relatively unwordlike items (such as *paije* and *soaz*, Seidenberg et al., 1996), and nevertheless uncovered phonological effects. Such findings lend weight to the strong phonology hypothesis; if unwordlike items activate base word phonology even when, in lexical decision, this is not helpful to the task in hand, this supports the view that phonological activation is automatic and mandatory (although, as was noted earlier, the processes driving the effect might be different for wordlike and unwordlike items).

Summary

Pseudohomophones are slower to respond to in lexical decision because of their phonological lexical status. However, these phonological representations are not frequency sensitive, but the links between phonological and orthographic representations are. Visual lexical decisions are affected by wordlikeness, such that more wordlike items take longer to respond to.

3.3 Experiment 4:
Phonological lexical decision with wordlike stimuli

Rationale

It has been argued that the phonological lexicon is not frequency sensitive, in opposition to the formulations of the influential DRC model and its near neighbour the MROM-P. One way to test this is to ask participants explicitly to access phonology, by using the phonological lexical decision task. If a base word frequency effect occurs for pseudohomophones when participants are asked whether a nonword sounds like a real word, then it can be concluded that the phonological lexicon is frequency sensitive. If it does not occur, then additional support is offered for the notion that the locus of the base word frequency effect must be sought elsewhere. The experiment also gives the opportunity to test the new more wordlike stimuli; are decisions in the phonological task also affected by a reader's orthotactic knowledge? In visual lexical decision, responses were slowed by increasing wordlikeness, but in naming, responses were speeded. Given that this task is more akin to naming, one might expect that more wordlike items will generate faster responses.

3.3.1 Method

Participants

33 students at the University of Bristol who had not taken part in the previous experiments participated in return for partial course credit.

Apparatus

The equipment set-up and display appearance and timing were the same as for Experiment 3, except that in this case the target letter string remained visible for 5 seconds.

Stimuli and design

The 52 pseudohomophones used in Experiment 3 were used as the target stimuli, together with 52 real words with a similar frequency range, and the 104 nonwords used in Experiments 1 and 2. (See Appendices A and B). Stimuli were randomly presented in blocks as in Experiment 3.

Procedure

The procedure was the same as for Experiment 3 except that participants were informed that their task was to judge whether the letter strings sounded like an English word or not.

3.3.2 Results

Errors were excluded from the reaction time analysis and trimming was carried out as in Experiment 3. Table 12 indicates the percentage of errors resulting from genuine errors and the trimming process; the error rate is noticeably high for pseudohomophones.

Table 12

Error rates (%) in the phonological lexical decision task.

	Pseudohomophones	Nonwords	Words
Before trimming	28.1	4.8	8.3
After trimming	29.8	5.7	12.4

Latencies for words were 314 ms faster than for pseudohomophones, which were 160 ms faster than nonwords (Table 13). Response times were longer than those in the previous experiment, and the direction of effect for pseudohomophones and control nonwords was different from Experiment 3 in that responses to pseudohomophones were faster than to nonwords (Figure 12).

Table 13

Mean response latencies and standard deviations for responses in phonological lexical decision.

	Mean RT (ms)	S.d.
Pseudohomophones	956	131
Control nonwords	1116	103
Words	642	74

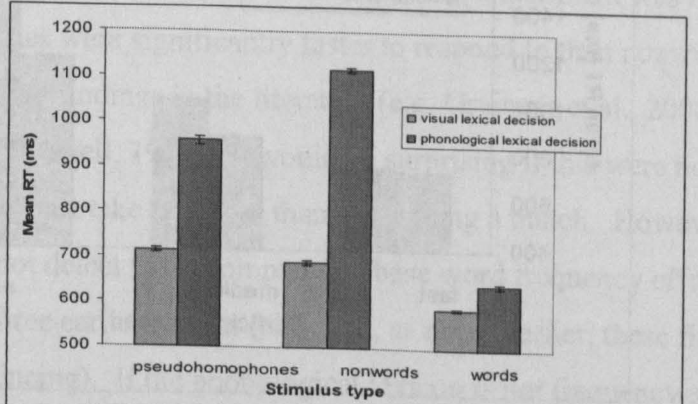


Figure 12. Mean response times and standard errors of the mean for Experiments 3 and 4.

As with the previous experiment, analyses of variance by subjects and by items followed by planned comparisons using t tests for the pseudohomophones and nonwords were carried out. With reaction time as the dependent variable, all tests were significant: F1 (2, 64) = 58.26, $p < .001$, F2 (2, 205) = 350.75, $p < .001$; t1 (32) = 4.65, $p < .001$, t2 (154) = 8.302, $p < .001$. With errors as the dependent variable, the ANOVAs were significant (F (2, 64) = 35.60, $p < .001$, F2 (2, 205) = 62.98, $p < .001$) while the t tests were significant by items but not by subjects (t1 (32) = 1.31, n.s., t2 (66.17) = 6.73, $p < .001$).

Reaction time to pseudohomophones and errors were correlated with base word frequency, N and wordlikeness. There were no significant correlations; the strongest relationship was between wordlikeness and reaction time (r (52) = - .24, $p < .1$).

Participant speed

Participants' responses were again divided into fast, medium and slow. Fast participants responded to pseudohomophones 14 ms faster than nonwords, for medium, the effect was 157 ms, and for slow, 399 ms (Figure 13). There was a significant main effect of group, F (2, 462) = 801.35, $p < .001$, and of nonword type, F(1,462) = 193.16, $p < .001$. The interaction was also significant, F (2, 462) = 67.39, $p < .001$. Planned t-tests indicated that the effect was significant only for the medium and slow participants: t (154) = 15.9, $p < .001$, t (154) = 12.43, $p < .001$.

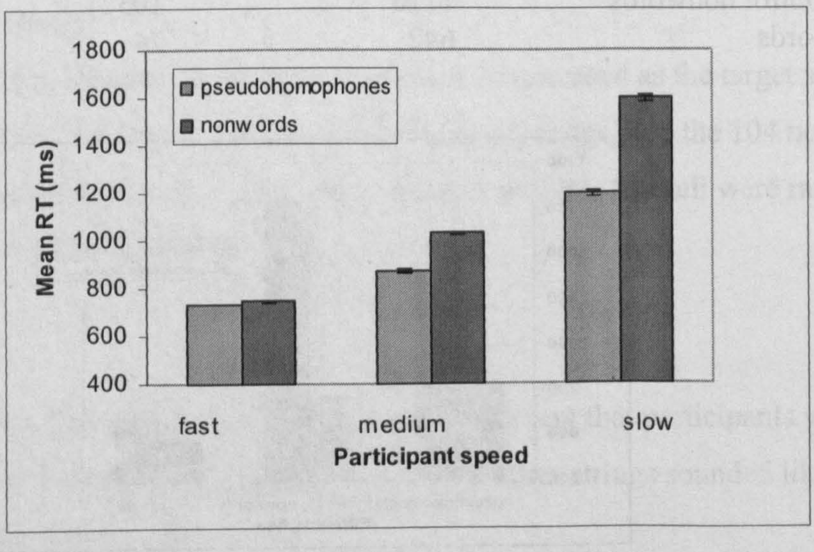


Figure 13. Pseudohomophone and nonword response times by participant speed.

Table 14
Percentage errors by participant speed.

	Pseudohomophones	Nonwords
Fast	28.1	10.7
Medium	37.2	11.1
Slow	23.9	15.2

Latencies for the three groups were correlated with base word $\log_{10}$ frequency, N, and wordlikeness; higher N items resulted in faster responses for the slow group (r (52) = -.30, p = .03) and there was a marginal correlation with wordlikeness for the fast group (r (52) = -.26, p = .06). Base word frequency did not approach significance for any group.

Table 14 shows that slow participants were more likely to make errors to nonwords than fast and medium readers, but the opposite pattern is apparent for pseudohomophones, where slow readers make fewest errors (F (2) = 3.75, p = .035). Pseudohomophone errors were correlated with frequency, N and wordlikeness. Neither N nor frequency correlated, but wordlikeness was significant for medium and marginal for slower ss (r (52) = -.297, p = .03, r (52) = -.267, p = .055), indicating that these participants tended to make more errors to the less wordlike items.

3.3.3 Discussion

The main result of the phonological lexical decision experiment was that pseudohomophones were significantly faster to respond to than nonwords, replicating standard findings in the literature (e.g. Grainger et al., 2000; McCann et al., 1988; Taft & Russell, 1992). It would be surprising if this were not so; finding a match to an item must take less time than not finding a match. However, the current experiment did not detect the accompanying base word frequency effect that was reported in the three earlier studies (although, as noted earlier, these findings may not be entirely convincing). If the phonological lexicon is not frequency-sensitive, then the absence of a base word frequency effect would be the expected outcome. The task explicitly asks participants to access phonology, and, by implication, to deemphasise the contribution from orthography; if, as argued in Experiment 3, frequency effects arise from the links between lexical phonology and orthography, then we should not expect to see frequency effects in the phonological lexical decision task, because orthographic representations are not being accessed.

A methodological difference should be commented on at this point; the earlier studies used lists containing nonwords and pseudohomophones, whereas the current experiment's lists also included real words. Therefore it is possible that participants in the two types of stimulus presentation were employing different strategies. When only nonwords and pseudohomophones are presented, finding a match to the input in the orthographic lexicon is a useful source of information; both phonology and base word spelling signal 'yes', even though the input signals 'no'. When words are also included in the list, matches in the orthographic lexicon are less informative; although they always result in a 'yes' response, sometimes the orthographic input conflicts with base word spelling and sometimes it does not. In this situation, the best strategy might be to ignore information from the orthographic lexicon; and if the orthographic lexicon is not activated, no frequency effects will occur.
This explanation might go some way towards accounting for the strikingly high pseudohomophone error rate in this experiment. Previous studies have reported fairly high error rates, but not of this magnitude (e.g. 15.9% in McCann et al.). Why did participants find it so difficult to correctly accept pseudohomophones? It cannot be that participants did not know the words in question; for, although we might accept that students do not recognise the phonological lexical status of *lirk* and *deam* (40%, 55% errors), it cannot be the case that they do not recognise *graid* and *hoam* (48%, 36% errors). It seems that, when the stimulus is a wordlike nonword, the phonological representation of the base word is not enough on its own to generate correct 'yes' responses. In earlier experiments, where orthography was more likely to be accessed, either where there were no words in the stimulus list, and/or where the stimuli were only one letter different from the base word, fewer errors were reported. In an experiment where all stimuli are wordlike, the combination of plausible orthographic input decoupled from lexical orthography results in a high error rate.

The position is further complicated by the fact that although pseudohomophone responses are more error-prone than those to nonwords, they are faster. This apparent speed-for-accuracy trade-off needs to be accounted for. If information from the orthographic lexicon is somewhat deactivated, as suggested above, this would mean that pseudohomophones will still generally find a (phonological) lexical match more quickly than nonwords, but weakened lexical orthographic information is more likely to prevent a successful match than when lexical orthography is also available as a strong source of information. Nonwords will not find a match, so will be slower, but in terms of accuracy they are advantaged, because a correct response does not implicate any activation of orthographic representations. (It might be argued that the participants did not understand the task. However, this seems unlikely for three reasons: first, the high academic profile of University of Bristol student participants; the clear nature of the experimental instructions; and a practice session which was carefully monitored for errors by the experimenter. Nevertheless, it is possible that, since visual lexical decision experiments are the more common paradigm in word recognition experiments, the participants were to some extent transferring knowledge of previous experiments to this relatively new scenario, that is, they were to some degree performing a visual lexical decision task).

The key point remains that that this apparently phonological task implicates orthographic processing. A similar observation was made by Taft (2006): "Here we have a task that would seem to be best carried out via the phonological system, yet orthographic information is unquestionably involved in making the response" (p. 84). This conclusion came from the observation of a high error rate in detecting the homophony of items like *seej*. Taft argued that this arose from the lack of orthographic overlap with the base word, and suggested that pseudohomophones with a greater degree of overlap would be more likely to activate correct responses in lexical decision. For example, *skream* will more successfully activate correct responses than *skreme* by virtue of its greater orthographic overlap with the base word *scream*. Taft argued that the processes that matched pseudohomophone input to base word orthography are via a non-phonological route, and that this evidence flew in the face of the claims of 'strong' phonology. The alternative view proposed here is that orthotactic probability rather than orthographic overlap is the important factor and that phonology is deeply implicated in processing orthotactically probable letter strings. Data from Experiment 4 allow a brief post hoc exploration of this view; the pseudohomophones that were identical to their base words apart from one letter (e.g. *ferm, sence*) were compared with those that were more than one letter different (e.g. *thort, hurse*). According to the orthographic match view, the former stimuli should generate fewer errors than the latter, but there was no significant difference in error rates (t2 (50) = 1.02, n.s.). Therefore it is arguable that the real reason why items like *seej* and *skreme* don't generate recognisable phonological lexical representations is not because they don't look like their base words, but because they don't look like any words. Rather than failing to make a direct orthographic match, they inhibit the activation of lexical phonology, and this prevents lexical orthographic access.

Summary

Experiment 4 showed that there is no base word frequency effect in latencies or errors under conditions when people are specifically asked to access the phonological lexicon. We must assume therefore that the phonological lexicon is not frequency sensitive. This experiment generated a very high number of errors, which may be attributable to the composition of the lists, and/or the wordlikeness of the stimuli. Visual lexical decision experiments give evidence for the automatic activation of phonology; this experiment supplies the corollary in that it offers evidence to suggest that phonological processing is closely linked to lexical orthographic knowledge. As such, it is additional evidence for a system that includes bi-directional feedforward and feedback mechanisms.

3.4 What is visual wordlikeness?

The rationale for creating more wordlike stimuli was to determine whether it was possible to detect a coherent pattern of responses to pseudohomophones in traditional naming and lexical decision tasks. Overall, Experiments 1 to 4 have shown that wordlikeness is an influential variable with effects that can be detected under these different experimental conditions. It speeded naming generally, and particularly under conditions where phonological processing was de-emphasised (that is, in mixed lists). In case alternated naming, a wordlikeness effect was still apparent when all other contributors to naming speed had been washed out. In lexical decision more wordlike items generated longer decision times and more errors. In phonological lexical decision, there was an effect of this orthographic variable even when the task required an emphasis on phonological processing. It is possible that the results of previous experiments have been confounded by the use of unwordlike items. The next step therefore is to establish whether the items used in previous studies were less wordlike than the current stimuli, and, if this is so, to establish what orthographic features contribute to a letter string being rated as more or less wordlike. Since the focus of this work is on pseudohomophones, the analysis concentrates on these items.

3.4.1 Wordlikeness analysis

387 pseudohomophones were originally rated for wordlikeness (see Section 2.2.1); these were the new items together with those used by previous researchers. Inspection of the stimuli led to 41 pseudohomophones being omitted from the analysis, either because English readers do not read them as pseudohomophonic (e.g. *bawx, rawk, stawp*), or because they are real words (e.g. *teem*). In addition, because the focus of this work has been on monosyllables, the 63 disyllabic items were removed from the analysis. Finally, 21 three-letter pseudohomophones were removed because there were relatively few of them, and it seemed likely that raters were processing these items differently from the longer nonwords, because they were all rated as very unwordlike (Table 15). Analysis showed that there was a marginal correlation between word length and wordlikeness such that longer items were apparently rated as more wordlike ($r(283) = .11$, $p = .06$), but this was accounted for entirely by the 3-letter items; when they were removed, there was no correlation between number of letters and wordlikeness ratings ($r(262) = -.06$, n.s.).

Table 15
Correlations between pseudohomophone length and mean wordlikeness.

		Mean wordlikeness	n
Length	3 letters	- .50	21
	4 letters	.20	130
	5 letters	.21	122
	6 letters	- .20	10

This left 262 pseudohomophone targets for analysis. (See Appendix C for ratings to previous researchers' stimuli).

Wordlikeness by stimulus set

One of the motivations for the wordlikeness rating exercise was to explore the notion that experimental stimuli used in earlier research were not wordlike, so the first analysis was to calculate the ratings given to the different stimulus sets. Table 16 shows that the new items, created from real words, were rated as more wordlike than the items created for previous research. A one-way ANOVA indicated significant mean differences ($F(10,251) = 8.07$, $p < .001$), with post hoc tests suggesting that although most of the previous stimuli did not generate ratings that were systematically significantly different from each other, they did differ significantly from the new stimuli used in experiments 1 – 4. Only two sets of items did not show a difference in wordlikeness from the new stimuli; these were Herdman et al.'s (1996) stimuli with legal rimes; and Taft and Russell's (1992) items. (See Appendix C for post hoc tests comparing the differences between wordlikeness ratings for the new items and previous stimuli).

Table 16
Mean wordlikeness ratings for pseudohomophones.

Mean z score	Stimulus set
0.70	New stimuli
0.65	Herdman et al. (1996) low frequency, legal rimes
0.46	Herdman et al. (1996) high frequency, legal rimes
0.31	Taft & Russell (1992) high frequency
0.20	Seidenberg et al (1996)
0.20	Taft & Russell (1992) low frequency
-0.07	Borowsky et al. (2002)
-0.05	Marmurek & Kwantes (1996) low frequency
-0.05	Marmurek & Kwantes (1996) high frequency
-0.17	Herdman et al. (1996) high frequency, illegal rimes
-0.30	Herdman et al. (1996) low frequency, illegal rimes

Wordlikeness, N and base word frequency

Raters were asked to rate the visual appearance of the pseudohomophones and to ignore their sound as far as possible, therefore a base word frequency effect would not necessarily be predicted. But, as we have seen in the lexical decision task, such an effect can occur even when the task is primarily orthographic. On the other hand, an effect of N would be predicted in that items with more neighbours would presumably be accorded higher ratings than items with fewer neighbours.

Both baseword frequency and N correlated with wordlikeness ratings (frequency, $r(262) = -.17$, $p = .006$; N, $r(262) = .52$, $p < .001$). The negative correlation with frequency indicates that raters gave higher ratings to less frequent items, and lower ratings to more frequent items, which is to be expected since readers' spelling knowledge is likely to be less secure for lower-frequency words. The positive correlation with N shows that items with more neighbours were given higher ratings. It has been argued that wordlike and unwordlike stimuli are likely to elicit different strategies from readers, and this notion was tested by comparing the effects of N and base word frequency with the most and the least wordlike stimuli. Correlations with the 100 most wordlike and the 100 least wordlike pseudohomophone stimuli showed that base word frequency was activated for the wordlike items, but not for the unwordlike items. The reverse effect occurred for N; it was not important for the wordlike items, but it was for the unwordlike (Table 17). This is additional support for the view that when orthotactic violations occur, lexical phonology is not activated. Readers instead focus on sources of orthographic information, such as N.

Table 17

Correlations for base word frequency and N with wordlikeness ratings.

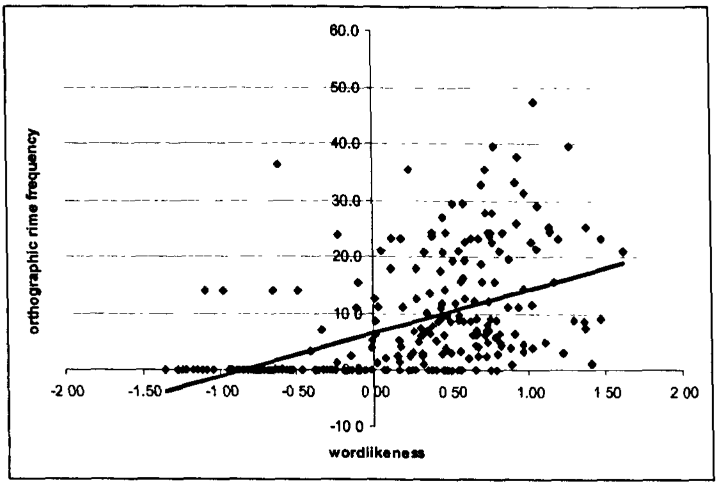
	Wordlike	Unwordlike
	r value	r value
Base word frequency	-.24*	.13 n.s.
N	.09, n.s.	.44**

** = p <.05, ** = p < .001*

What is the nature of the orthotactic violations that prevent lexical access? One potential candidate is the rime; indeed, Ziegler and Perry (1998) have suggested that

N is less a measure of spelling neighbours than an index of orthographic rime frequency, since most neighbours in English are rime neighbours. Further, as can be seen in Table 16, those stimuli termed 'legal' by Herdman et al. on the basis of rime spelling were accorded higher ratings than those stimuli with 'illegal' rimes. Inspection of the stimuli suggested that pseudohomophones with more frequently occurring rimes (e.g. *-ane, -oar*) were given higher ratings than those with non-existent or very low frequency rimes (e.g. *-ene, -ayn*).

With rime frequency calculated as the summed $\log_{10}$ frequency of all monosyllabic words containing that rime spelling, the mean wordlikeness rating for low-frequency (< 3) or non-existing rimes was -.25; for rimes with a frequency of between 4 and 47.5, the mean wordlikeness rating was .57, a significant difference ($t(260) = 16.18$, $p < .001$). The correlation between orthographic rime frequency and wordlikeness rating was also significant, $r(262) = .50$, $p < .001$. The strength of the correlation weakens as rimes become more frequent, so that for the 50 items with the most frequent rimes, rime spelling was no longer a predictor of wordikeness ($r(50) = .109$, $p = .45$). See also Figure 14 below.



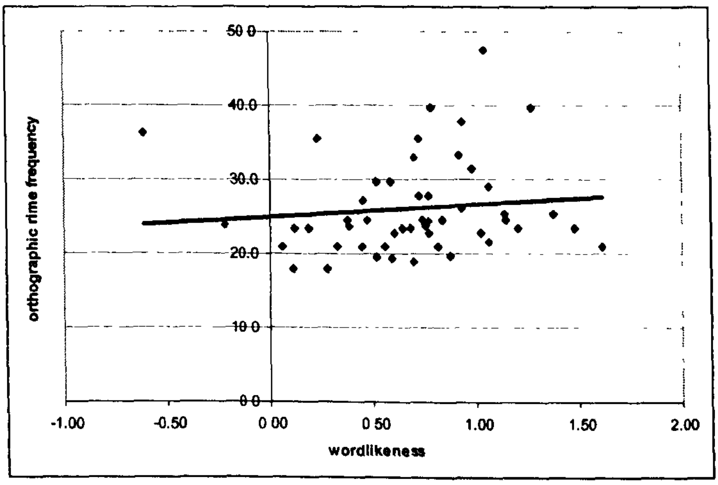


Figure 14. Including non-existent and low-frequency orthographic rimes (top) affects wordlikeness ratings more than existent, higher-frequency rimes (bottom).

3.5 Summary and conclusion

Readers respond to the wordlikeness of letter strings; this was shown in the experiments reported in this chapter and in the analysis of the wordlikeness ratings. Evidence from using more wordlike stimuli offers support for the notion of an interactive reading system with bidirectional feedforward and feedback connections. Wordlike pseudohomophones activate phonological and orthographic lexical representations, although the nature of this activation is constrained by the task's requirements. Unwordlike stimuli are less likely to activate phonological lexical representations and it is therefore unsurprising that experiments using unwordlike stimuli have found null or unclear phonological effects. Orthographic rime spelling seems to be a primary contributor towards a reader's perception of wordlikeness, with non-existent or very low frequency rimes generating low wordlikeness ratings.

Chapter 4

Is Rime the Reason?

"I take it you already know
Of tough and bough and cough and dough?
Some may stumble, but not you,
On hiccough, thorough, slough, and through."
(Adams, 1990)

This chapter considers the possibility that readers can not only detect spellings that are highly improbable in English spelling but that they can also detect patterns within degrees of orthotactic possibilities. If readers develop a sensitivity to statistical spelling patterns, then it is likely that they are not only sensitive to the fact that *-ayn* is a highly improbable spelling for the phonology /eɪn/, they are also sensitive to the fact that *–ane* and *–ain* are both potential candidates, and, further, that the latter is a more likely occurrence than the former because it occurs more frequently in written English. Such evidence would offer support to the view that responses to nonword letter strings are affected by readers' implicit orthotactic knowledge, and would also offer additional information as to the likely nature of the reading system itself. In Chapter 3 it was argued that feedback from orthographic lexical representations was an important contributor to lexical decisions; even in the phonological task, where such information is unhelpful, and it is likely that information from the rime was a likely contributor to feedback effects. Rimes are particularly useful to investigate, because they are a strong source of variance in English spelling, playing "a special role in the description, learning, and use of the English writing system" (Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995, p. 108). As we have already seen, rime spelling is an important contributor to wordlikeness ratings. It would be expected that items with less common rime spellings would be more likely to result in errors and slower naming and lexical decision times, since secure orthographic representations would be less susceptible to conflict than weaker representations.

4.1 The status of rimes in reading

The previous chapters have shown that readers can consistently rate letter strings as more or less wordlike, and that orthographic wordlikeness affects responses in a number of tasks typically used to uncover the processes involved in reading. It has been argued that wordlikeness is a variable that may have contributed to the inconclusive results reported in the literature. It is important to uncover the contributory components of wordlikeness, and to establish what role they may play in experimental tasks. One likely candidate that has been identified is the rime element of the stimulus. We have seen that items with non-existent or very low frequency rimes were rated as very unwordlike; a clear demonstration of this was given by the different ratings given to Herdman et al.'s (1996) legal and illegal rime stimuli. The question now arises as to whether readers demonstrate sensitivity to the relative frequencies with which legitimate rimes occur. Some indication that this is likely is offered by Ziegler and Perry (1998), who found that words with frequent rimes (such as *clan* and *trash*) were responded to more quickly in lexical decision than words with infrequent rimes (such as *curd* and *badge*). In addition, so-called 'strange' (or 'exception') words, such as *guard, weird* and *debt*, have long been known to elicit slower decisions in lexical decision (Waters & Seidenberg, 1985); their uncommon spellings are predominantly attributable to unusual rimes. Rosson (1985) found that adults took longer and made more errors in naming nonwords with no orthographic rime neighbours than nonwords with at least one orthographic rime neighbour. Brown (1987) found that participants named consistent regular words containing high frequency orthographic rimes faster than words with unique orthographic rimes.

The argument for investigating the role of rimes is supported by previous work, which has shown that rimes affect responses in various experimental tasks. Glushko (1979) initially established that regular consistent words such as *wade* were named more quickly than regular but inconsistent words such as *wave*, by virtue of conflict with alternative pronunciations such as in *have*. This was confirmed by Andrews (1982), not only in naming tasks, which explicitly require computation of a phonological code, but also in visual lexical decision. These findings extend also to nonwords, such that rime consistency influences nonword reading time (Glushko,

1979) and inconsistent nonwords show a higher rate of irregular pronunciations than consistent nonwords (Andrews & Scarratt, 1998; Glushko, 1979).

There is a wide range of evidence that suggests that orthographic rimes function as units of print for adults (e.g. Bowey, 1990, 1993; Taraban & McClelland, 1987; Treiman, 1984; Trieman & Chafetz, 1987; Treiman & Zukowski, 1988). Evidence that readers base their pronunciations on experience with previous orthographic rimes was reported by Treiman, Goswami and Bruck (1990) who found that people were better at pronouncing nonwords that shared rimes with many real words (e.g. *goach*, *tain*) than nonwords that shared rimes with fewer real words (e.g. *taich*, *boan*). Rime is important and accounts for more variance in naming latency than any other orthographic unit (Treiman et al. 1995) because it contributes to reducing ambiguity in pronouncing vowels, which are the most inconsistent elements of English spelling-sound correspondences. These authors argue that rime units are a better source of correct pronunciation information than onset plus vowel units because the final consonant is often a good guide to the correct pronunciation of the vowel. For example, although *ea* can be /ɛ/ or /i/ as in *head* or *bead*, *-eap* is nearly always /ip/. By contrast, the initial consonant plus vowel is not usually a particularly useful pointer towards the correct pronunciation of the vowel.

Previous research has largely focused on feedforward consistency effects in rimes; that is, where the same spelling leads to different phonologies. By contrast, all the items used in Experiments 1 - 4 were, because of the stimulus construction procedure, of necessity feedback inconsistent. The number of studies exploring feedback inconsistent stimuli has been limited until relatively recently, possibly because the prevailing paradigm in the literature has been that of an essentially unidirectional, feedforward, mapping of spelling to sound. The following studies, however, did present results based on feedback inconsistent stimuli. Stone et al. (1997) found that words with rimes that could be spelled more than one way (e.g. -*eap* and *-eep* for /ip/) were slower to elicit correct 'yes' responses in visual lexical decision than words with rimes that could only be spelled one way (e.g. *–obe*). They did not investigate whether more frequently occurring spellings of the same phonological rime elicited faster responses than less frequently occurring spellings. Ziegler et al. (1997) replicated Stone et al.'s work (in French) and found evidence for feedback effects in both naming and visual lexical decision. However, neither Stone et al. nor Ziegler et al. uncovered consistency effects for nonwords.

Further evidence for a feedback effect was offered by Pexman et al. (2001) and Pexman, Lupker, and Reggin (2002) who reported slower responses in naming and visual lexical decision for feedback inconsistent items and suggested that these were generated by competing activations in orthographic units. They interpret their findings in terms of a general connectionist-type model, but acknowledge that the dual-route approach could also explain the findings in terms of feedback from phonology activating competing lexical entries. The issue of whether feedback effects occur from lexical or sublexical processes is unresolved; Pexman and colleagues came to the conclusion that the effect was probably lexical, while Stone et al. argued for effects at the rime level, and Ziegler et al. proposed two mechanisms, one lexical and one sublexical. None of the above studies investigated feedback inconsistent nonwords and pseudohomophones, of the type that have been used in Experiments 1 – 4.

Finally, support for the view that readers will show sensitivity to the relative frequency of rimes comes from work by Ziegler et al. (2001, Expt. 3), who showed that pseudohomophones and nonwords derived from German base words with dominant spelling patterns were quicker to reject in visual lexical decision than those derived from base words with subdominant spelling patterns. A more stringent test of the hypothesis that bi-directional orthography-phonology mappings are activated would be if evidence emerged for a rime effect in phonological lexical decision. In phonological lexical decision, the emphasis is explicitly on phonology, so a response can theoretically be made purely from activation in the phonological lexicon. If feedback nevertheless occurs that is attributable to orthographic activation via phonology, then it would be expected that well-entrenched or more stable spellings of dominant rimes will conflict with the less stable inputs of subdominant rimes, and therefore subdominant rimes will be slower to elicit 'yes' responses. Of course, it could be argued that a rime effect can occur without access to the lexical or sublexical orthography, and that subdominant rimes are simply slower to compute, so a rime dominance effect will occur in any case. However, if this is the case, then it is to be expected that nonwords will also show an effect of rime dominance, so that both pseudohomophones and nonwords with dominant rimes will elicit faster responses.

Rimes in the wordlikeness ratings

If readers are sensitive to the relative frequencies of alternative orthographies for phonological rimes, then evidence might be expected to emerge from the wordlikeness ratings. A *post hoc* inspection unfortunately only offered a few cases where there were two ways of spelling the same real word phonology; but in this limited set it was clear that different ratings were given to different orthographies. As we have seen, non-existent rimes were given very low ratings, and higher ratings were given for existing rimes. Where both rimes exist in English spelling, as for *lerk* and *lirk*, higher ratings were given for both; but note that *lerk* is rated as more wordlike than *lirk.*

Table 18

Ratings for alternative pseudohomophone rime spellings.

PH	Wordlikeness (mean z score)	PH	Wordlikeness (mean z score)
chane	0.64	*chayn*	-0.79
pruve	0.09	*proov*	-0.62
traid	1.17	*trayd*	-0.76
lerk	0.97	*lirk*	0.67

The question now is whether existing but different rime spellings are responded to on the likelihood of their probabilities. Experiment 4 offered some encouraging indications that this might be the case, in that subdominant rimes generated more errors and were responded to more slowly than dominant rimes (Table 19).

Table 19

Latencies and errors for dominant and subdominant rimes in Experiment 4 (phonological lexical decision).

	Mean errors	Mean RT (ms)
Dominant rimes (n = 17)	8.9	975
Equal rimes (n = 14)	7.8	989
Subdominant rimes (n = 21)	10.6	1010

However, the difference was non-significant, and the rimes were not directly comparable because each set contained different items. There were three pairs of items where direct comparisons could be made, and it can be seen that more errors were made to the pseudohomophones with dominant orthographic rimes (Table 20).

Table 20
Errors for three comparable pseudohomophones by rime status from Experiment 4 (phonological lexical decision).

Dominant rime	Errors	Subdominant rime	Errors
crain	2	*chane*	12
soke	3	*spoak*	8
sope	4	*poap*	16

On the basis of these data it would seem worthwhile to manipulate rime status *a priori*, in order to investigate more systematically the view that readers are sensitive to probabilistic rime patterns.

How might models account for the role of rimes?

Another reason for exploring rime status is that theoretical and computational accounts of reading make differing predictions. The suggestion so far is that readers detect regularities in the orthography of their native language, and that this knowledge transfers to processes involved in naming and visual lexical decision tasks. Thus, not only do readers detect a coarse-grained difference between existent and non-existent rimes, such as *–ain* and *–ayn*, they also show a finer-grained sensitivity to the differences between more or less frequently occurring rimes, such as *–ain* and *–ane*. Few models incorporate rime explicitly in their models; the exceptions are Plaut et al. (1996), and Zorzi et al. (1998).

The DRC model

In theory, rules for rimes could be something that the reading system learns and would therefore be part of the nonlexical route. Patterson and Morton (1985) proposed that the sublexical route consisted of a grapheme-phoneme correspondence route and a "body" system that conveyed information about rime constituents. However, in the computational dual-route cascaded model (Coltheart et al., 2001) rimes are not represented in the GPC route because the rules are restricted to cases where a letter, or set of letters, maps on to a single phoneme. Thus, the model is likely to deal with existing and non-existing rimes in exactly the same way; and indeed, a quick simulation shows that *wirds, werds,* and *wyrdz* are all named in exactly the same number of cycles (511). Rime effects might emerge from the lexical route through the activation of similarly-spelled words, so that *rane* would activate *lane, mane, pane* etc in the orthographic lexicon while *sain* would activate *gain, main, vain,* and *rain.* However, these inputs would also activate other, non-rime-related neighbours, such as *rant, rage, rate/sail, shin, spin* and it therefore seems unlikely that a rime dominance effect would emerge from these multiple activations. In any case, in the implemented model, lexical inhibition is built in so as to prevent lexicalizations of nonwords, thus nonwords tend to be read very effectively by the nonlexical route alone, i.e. without support from the lexical route.

PDP models

Sensitivity to rimes might be expected to emerge from PDP models, because they extract statistical regularities during the learning processes, and if rimes are useful guides to correct output, then this knowledge will develop. For example, Plaut et al. (1996) analysed the structure of the componential attractors that resulted from learning, and found "a slight interdependence among the vowel and coda, consistent with the fact that the word bodies capture important information in pronunciation" (p. 87). If feedback inconsistent rimes are included in the training corpus of a connectionist model, then it would be expected that the model would show different responses to these types of stimuli, so that, for example, less frequent rimes would be responded to more slowly. Also, it is likely that, if trained on a corpus of normal English words, such models are likely to produce inaccurate responses to strange combinations of letters. Indeed, Seidenberg and McClelland (1990) countered accusations that their 1989 model could not read nonwords very well by pointing out that the difficulty arose because the model had not been exposed to unusual rimes and was therefore unable to deal with stimuli such as *faije* and *vawx.* These are two of the items that we have now established as being unwordlike, apparently because of the non-existence of their rimes.

Interactive-activation/resonance models

Of most interest in terms of the theoretical approach that best suits the data so far is an interactive-activation or resonance type model, with bi-directional connections between orthographic and phonological units, at lexical or sublexical levels. The various models of this type fall into two groups, based on whether the processing is considered to be lexical or sublexical, and they generate different predictions for words, nonwords and pseudohomophones. Lexical matching models (e.g. McClelland & Rumelhart, 1981; Jacobs & Grainger, 1992; Paap et al., 1987) do not predict that rime dominance will affect processing for pseudohomophones and nonwords. For example, the phonology /hip/ would activate *heap,* but not *heep* since it does not exist in the orthographic lexicon. Lexical matching models therefore do not predict rime effects; input *rane* might activate the orthography *rain* but the status of the spelling of /eɪn/ is irrelevant. And since neither *frain* nor *frane* activates an orthographic representation, each will be responded to in the same time. Sublexical matching models (e.g. Grossberg & Stone, 1986; Stone & Van Orden, 1994; Van Orden & Goldinger, 1994) implement multiple sublexical components of different grain sizes; a rime pronounced as /-ip/ would activate two possible spellings, *-eep* and *–eap*. Therefore these models would be expected to demonstrate an effect of rime dominance, even in nonword processing, so *frain,* for example, will be processed differently from *frane* because *-ain* is the more frequently-occurring orthography. More recent models (e.g. Jacobs et al., 1998; Ziegler, Muneaux & Grainger, 2003) posit a central interface between orthography and phonology which allows sublexical orthographic representations to be mapped on to their corresponding phonological representations, which can influence the course of visual word recognition via their interaction with sublexical orthographic representations, or via the activation of whole word phonological representations. The focus of both types of model has been on word recognition, with predictions regarding nonwords and pseudohomophones usually left unspecified. It is of interest therefore to investigate the role of rime dominance in nonwords and pseudohomophones because this can shed further light on the nature of the relationships between lexical and sublexical processing. Very little research has been carried out in the area, and is contradictory; Stone et al. (1997) found no difference in response times to nonwords with feedback consistent and inconsistent rimes; but Ziegler et al. (2001) reported a spelling consistency effect for both pseudohomophones and nonwords.

The most stringent test of the view that there are bi-directional links between orthographic and phonological units is with another phonological lexical decision experiment, since, if effects of orthographic feedback can be detected, it is clear evidence for automatic activation of orthography even when it is actually detrimental to the task. The only experiment that has investigated feedback effects in phonological lexical decision was reported by Pexman et al. (2002, Expt. 2), who concluded that the task causes phonological processing to override orthographic activation; they reported that the orthographic feedback effects found in visual lexical decision disappeared in the phonological form of the task. If this is the case, a feedback effect based on rime orthography is unlikely; and both nonwords and pseudohomophones should show the same effects as far as the rime is concerned. On the other hand, if pseudohomophones and nonwords offer different patterns of rime results, then the conclusion would be that base word orthography was activated for the pseudohomophones. In addition, if evidence emerges that orthography is activated even when the task is apparently primarily phonological, it would, like Experiment 4, provide a neat counterbalance to the large number of experiments that have demonstrated that phonology is activated even when the task is apparently primarily orthographic, and as such, strong support for a model involving bi-directional connections.

Summary

Previous research has shown that the orthographic rime has an important role to play in tasks involving reading. Wordlikeness judgements are based, at least in part, on rime frequency. Therefore it is logical to explore the role of rime explicitly, and one way that this can be done is by investigating dominant and subdominant rimes. By definition, these rimes are feedback inconsistent, so it makes sense to locate this set of experiments within the context of theories and models of word recognition that include scope for feedforward and feedback connections between orthography and phonology.

4.2 Experiment 5

Phonological lexical decision using stimuli with dominant and subdominant rimes

Predictions

On the basis of the material described above, it is possible that 'yes' decisions to pseudohomophones with dominant rimes will be faster in a phonological lexical decision task, and fewer errors will be made to these items than to those with subdominant rimes. Such an effect is unlikely to occur for nonwords, because they do not activate lexical phonology. However, on the basis of Pexman et al.'s (2002) experiment, an effect attributable to feedback from orthography will be difficult to establish, since the phonological nature of the task weakens orthographic processes.

4.2.1 Method

Participants

37 University of Bristol students participated in the experiment in return for partial course credit.

Apparatus

The equipment set-up and display appearance and timing were the same as for Experiment 3.

Stimuli and design

The stimuli consisted of 62 words, 62 pseudohomophones and 124 nonwords; all were monosyllabic and between 4 and 7 letters. Pseudohomophone stimuli were created as described in Chapter 2, by crossing onsets and rimes of existing words, but this time rime dominance was also controlled so that there were equal numbers of stimuli with dominant or subdominant rimes. Rime dominance was calculated by summing the $\log_{10}$ frequency of words with that rime spelling. The base word frequency range was the same for dominant and subdominant rimes (Figure 15).

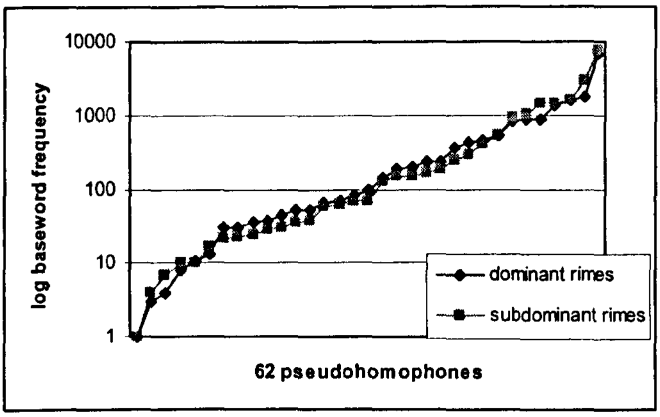


Figure 15. Equivalent base word frequency for stimuli with dominant and subdominant rimes.

The pseudohomophones were specifically cross-matched in terms of rime (e.g. *clame/taim*) whereas the words were not, in order to give a wider variety of orthographic rimes in the stimulus set. The nonwords were created by crossing the onsets and rimes of, first, the pseudohomophone and, second, the word stimuli. The mean rime frequencies for the dominant/subdominant rimes were 11.2/2.6 for words, 10.9/4.5 for pseudohomophones, and 11/3.6 for nonwords. The mean word frequencies for dominant/subdominant words were 946.8 and 965, and 346.4 and 392 for the base words of the pseudohomophones. These differences were not significant, for words, t (60) = -.042, n.s.; for pseudohomophones, t (60) = -.30, n.s. (See Appendix D for stimulus lists).

Procedure

The procedure was the same as for Experiment 4; stimuli were presented in quasi random order, and participants were asked to judge whether the letter string sounded like a word or not.

4.2.2 Results

Data from two participants who made more than 50% errors on the pseudohomophones were excluded from the analysis. Over half of the remaining 35 participants did not recognize the homophonic nature of the following: *deam, halk, spight, wafe, breem, glaid, teek*, so these items were removed from the analysis. Two real words, *lithe* and *kirk*, were also removed from the analysis because more than half the participants did not recognize their lexical status. For the remaining items, trials were excluded from the latency analysis if an incorrect response was made, or if the reaction time for the trial was more than 3.00 standard deviations from the participant's mean reaction time for that condition. It is worth noting that in spite of these procedures, there was considerable variability in response times, with a range from 526 ms to 3298 ms for all items taken together, with a mean of 1038 ms, but there are no participants or items that can be clearly defined as outliers and all three conditions show normal distributions, so no further trimming was carried out.

Since the nonwords were created from two different sources, the words and the pseudohomophones, an initial check was carried out to establish that there was no difference between them. A one-way ANOVA showed that there was no difference between reaction times to the two types of stimuli: $F(3, 120) = .58$, n.s. Therefore the two stimulus subsets are treated as a homogeneous group, labelled 'nonwords', in the analyses that follow.

As with Experiment 4, latencies to words, pseudohomophones and nonwords were shown to be fastest to words, slower to pseudohomophones, and slowest to nonwords (Table 21); this was significant: $F1\ (2,207) = 45.74$, $p < .001$; $F2\ (2,236) = 426.45$, $p < .001$.

Table 21

Mean response latencies, standard deviations, and percentage errors for responses in phonological lexical decision.

	Mean RT (ms)	S.d.	Errors (%)
Words	748	180	2.2
Pseudohomophones	989	129	6.8
Nonwords	1384	147	5.1

Responses by rime type also suggested that dominant rimes were faster to respond to than subdominant (Table 22).

Table 22
Mean decision latencies (ms) in phonological lexical decision by rime type.

	Dominant rime Mean RT *(S.d.)*	Subdominant rime Mean RT *(S.d.)*	Diff
Words	706 *(159)*	788 *(201)*	-82
Pseudohomophones	959 *(138)*	1019 *(120)*	-60
Nonwords	1375 *(140)*	1394 *(154)*	-19

Planned comparisons using t tests were used to investigate differences between latencies to different rime types. By participants, there were significant differences for words and pseudohomophones, but not for nonwords. By items, there was a significant difference for rime type in words and the difference for pseudohomophones approached significance. There was no difference in rime type for nonwords.

Table 23
t tests for differences between dominant and subdominant rimes in words, pseudohomophones, and nonwords.

	By participants	By items
Words	$t\ (34) = 4.32, p < .001$	$t\ (58) = 2.57, p = .013$
Pseudohomophones	$t\ (34) = 3.02, p = .005$	$t\ (53) = 1.80, p < .1$
Nonwords	$t\ (34) = .86$, n.s.	$t\ (122) = .68$, n.s.

The lexical status of the stimulus appears to affect responses; there is no difference in rime type for the nonwords, but there is for words and pseudohomophones. The by items analysis for pseudohomophones however, is non-significant, but further analysis indicates that a small number of items eliciting the slowest latencies might have been responsible for this. Figure 16 indicates that all the rime-dominant pseudohomophones are responded to faster than the subdominant stimuli until a point around 1100 ms, at which time the pattern of differences becomes less distinct. Removing the slowest five items from each stimulus set established that the difference between means for the pseudohomophones below this cut-off point was significant: $t\ (43) = -2.47, p = .017$.

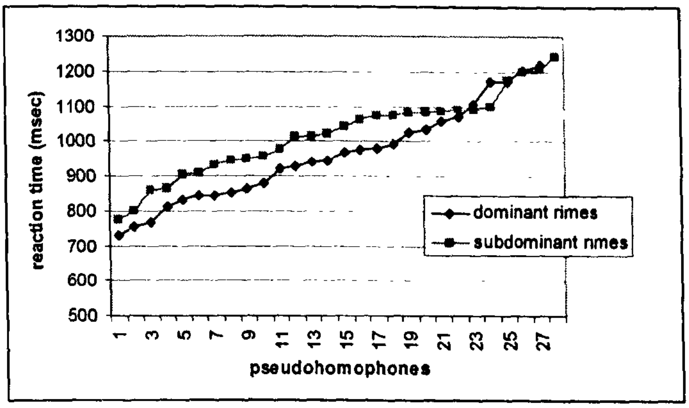


Figure 16. Pseudohomophone reaction time by orthographic rime type.

Error analysis

The above analysis was repeated using errors as the dependent variable. Words elicited fewer errors than pseudohomophones and nonwords (analysis by items, $F(2,236) = 12.92$, $p < .001$). All three types of stimuli with subdominant rimes generated more errors than those with dominant rimes (Table 24).

Table 24
Mean errors for words, pseudohomophones and nonwords

	All items	Dominant rime	Subdominant rime
Words	7	6	8
Pseudohomophones	18	15	20
Nonwords	17	16	18

By items analysis shows that none of these differences are significant: Words, $F(58) = .62$, n.s., pseudohomophones $F(53) = 1.59$, n.s., nonwords, $F(91) = .39$, n.s. In the by participants analysis, there are significant differences: Words $t(34) = 2.53$, $p = .016$, pseudohomophones $t(34) = 3.65$, $p <.001$, nonwords $t(34) = 2.36$, $p = .02$).

N and base word frequency

Latencies and errors to the pseudohomophones were correlated with $\log_{10}$ base word frequency, N, and, the metric of interest in this experiment, orthographic rime frequency. Correlations were not significant, although the correlation with reaction time and rime frequency gave $r(55) = -.24$, $p = .076$. However, when the lowestfrequency rimes were removed from the analysis, even this weak association disappeared: r (46) = .01, n.s.

Summary

The usual pattern of responses in phonological lexical decision occurred, in that words elicited fastest responses, followed by pseudohomophones and finally by words. In addition, dominant rimes facilitated decision latencies to words and pseudohomophones over subdominant rimes, but there was no difference for nonwords by rime type. As with the previous phonological lexical decision experiment, there was no effect of base word frequency.

4.2.3 Discussion

As with Experiment 4, this phonological lexical decision task resulted in slower responses to nonwords than to pseudohomophones, but again there was no base word frequency effect for the pseudohomophones. In addition, pseudohomophone responses were affected by rime status such that more frequent rime spellings resulted in faster responses; this mirrored the findings with words. These findings offer further support for the notion that pseudohomophone processing is affected by the wordlikeness of the stimulus; we have already seen that items with frequently occurring rimes tend to be given higher wordlikeness ratings than those with non-existent or very low frequency rimes. Readers are not only sensitive to spellings that can be characterised as part of a legal/illegal or existent/non-existent dichotomy, they are also sensitive to the relative frequency with which a phonological rime can be represented orthographically. Expressing this sensitivity as a probabilistic continuum is likely to be a better way of representing readers' knowledge of spelling, running from highly unlikely through to possible but not probable to highly probable. This is also likely to capture individual differences in reader knowledge and may therefore offer more insights than a one-size-fits-all binary rule system.

The experiment has indicated that rime frequency affects word and pseudohomophone processing, but not nonwords, and, if this is the case, then this has interesting implications for the word recognition system, because it implies that the lexical status of pseudohomophones leads to different orthographic processes than those used for nonwords. Further, it also implies that the processes underpinning phonological lexical decision are not, as is frequently claimed, solely reliant on phonology, because this experiment has given clear orthographic effects. How confident can we be that nonwords genuinely do not elicit the same results as for words and pseudohomophones? Previous research is limited to an experiment in German, where an effect of spelling dominance for both pseudohomophones and nonwords was reported (Ziegler et al., 2001). However, the spelling probability effect was not significant in the by items analysis, which treated pseudohomophones and nonwords together. Stone et al. (1997) found no difference in responses to true nonwords with consistent and inconsistent rime spellings. Taken together, these two findings suggest that nonwords may not elicit the same responses as pseudohomophones. However, the general pattern (Table 22) was for nonwords to elicit faster responses for the dominant items, so it is possible that a type 2 error has occurred in the above analysis.

If we assume that nonwords did not show the same effects of rime dominance as words and pseudohomophones, we can argue that sensitivity to rime spelling appears to result, in this task at least, from lexical knowledge. In this task, words and pseudohomophones both activate phonological units, which in turn activate orthographic representations. Nonwords, such as *frain* and *frane,* do not activate any real word orthographic representations, and therefore there is no difference in response time for the two spellings. Both letter-strings *brain* and *brane* activate the phonology /breɪn/, and two things might happen as a consequence of this activation. One is that the spelling of *brain* is activated. In addition, other words with the same-sounding rime, such as *train, drain, lane* and *sane* are activated. Among these, the rime *-ain* is dominant, so there will be more activations of words with this orthographic rime, and therefore any input with a dominant rime will be processed faster than any input with a subdominant rime, regardless of whether it is a real word or a pseudohomophone. Any of the models described above that incorporate feedback and feedforward lexical mechanisms would be able to accommodate these results. Thus the apparently sublexical rime effect emerges from lexical activation. The other thing that might occur once lexical phonology has been activated is that the whole word spelling of *brain* is not activated so much as its sublexical orthographic constituents. So, the input *brane* activates the sound /breɪn/ which is decomposed so that the /eɪn/ phonology activates the (dominant) orthography *–ain,* which conflicts with the orthographic input. On the other hand, while the input *sain* also activates the /eɪn/ phonology, it also activates the dominant spelling which is identical to that of the input. So the decision time is longer for *brane* than *sain*. Given that there were no frequency effects, and the argument so far has been that frequency effects are likely to be located in the links between phonology and lexical orthography, then a sublexical activation account would be the preferred explanation. This view is supported by the fact that there was no significant correlation with base word frequency, but there was an effect of orthographic rime frequency (although this was attributable to the very low frequency items). Lexical and sublexical processes are not mutually exclusive; in the recent version of the interactive-activation models proposed by Grainger, Muneaux, Farioli, and Ziegler (2005), stimuli rapidly activate a set of sublexical phonological representations that interact with sublexical orthographic representations and whole word phonological representations.

An alternative interpretation might be that what appear to be feedback effects emerge from prelexical processing. Dominant rimes are simply computed more quickly than subdominant because they are encountered more often. If this were the case, nonwords should also show facilitation for dominant rimes, but, in this experiment at least, there is no sign of this.

What about errors?

Error analysis might help shed light on the locus of the rime effect, although it needs to be borne in mind that although the by subjects analysis was significant, the by items analysis was not. An argument could be made to support the view that errors occur at a prelexical encoding stage, since, overall, there were more errors for items with subdominant rimes than dominant, which suggests that errors arise from early prelexical encoding. However, if the rime effect for pseudohomophones occurs because of the activation of lexical or sublexical orthography, then it would be predicted that more errors would be made to pseudohomophones with *subdominant* rimes; that is, the input orthography conflicts with the dominant spelling, activated via phonology. Overall, that is what the error data suggest.

It should be noted in passing that the error rate in this phonological lexical decision experiment is much closer to those reported in previous research than the error rate in Experiment 4. It was suggested that the wordlikeness of the stimuli, or list composition, might have resulted in that experiment's high error rate, but Experiment 5 was designed in a very similar way; so the high error rate of Experiment 4 remains something of a mystery. The response times for Experiment 5 were on average 136 ms slower than those for Experiment 4, so it is possible that the speed-for-accuracy trade-off seen for pseudohomophones in that experiment is attributable to participants setting an unfeasibly fast response time criterion; the overall slower responses seen in Experiment 5 offered more time for phonological activation to become available for a conscious judgement to be made.

Summary

Participants are sensitive to the relative frequency with which phonological rimes are spelled, and showed this by responding differently to dominant and subdominant rime spellings in words and pseudohomophones. This sensitivity might occur at a prelexical stage, but, since nonwords did not show the rime dominance effect, the preferred interpretation is in terms of activation of lexical phonology spreading to orthographic representations; knowledge of the relative frequency of rime spellings is encoded here.

4.3 Experiment 6

Disrupting the rime by letter transposition

Aims and predictions

The following experiment was designed to explore the view that rime effects occur at a very early, prelexical stage. We have noted that rime generally is an important source of information in decoding words (e.g. Treiman et al., 1995) and more specifically we have seen that it affects wordlikeness judgements to pseudohomophones and nonwords, and responses in phonological lexical decision. If the reading system learns to recruit whatever regularities exist in monosyllabic words, it is possible that one useful piece of information would be about the rime-onset structure. In an early study, Treiman and Chafetz (1987; see also Treiman, 1994) found that adults made faster lexical decisions to words presented with slashes inserted between orthographic onset and rime units (e.g. CR//ISP) than to words with slashes inserted within the orthographic rime units (e.g. CRI//SP). It may be that the word recognition system has through experience learned to carry out an initial parsing of the input letters into onset and rime components (Figure 17). This parsing would occur after letters had been assigned to their relative position in the letter string, and would supply additional information as to the potential decoding of the letter string; in particular the first vowel in the letter string would signal the start of the rime. (Note that this scheme is specific to CVC monosyllables; it does not solve the problem of how vowel-initial units are encoded. For the purposes of the experiments reported here, which are solely concerned with CVC stimuli, this scheme is adequate, but clearly a further modification would be needed in order to deal with monosyllables that start with a vowel).

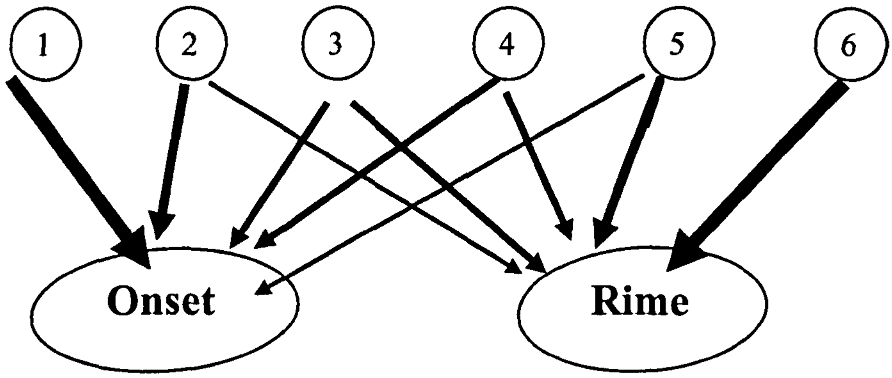


Figure 17. How letter positions might map on to onset and rime, prior to lexical coding. Connections between the first and last letters must be fixed, because they can only be associated with either onset or rime. The relative strength of connections with the internal letters will vary depending on their position and the number of letters in the word.

One way to establish whether prelexical processing is responsible for rime effects is to use a masked presentation technique that allows access to very early responses, as described in Frankish and Turner (2007). In order to manipulate the rime, the transposed letter (TL) technique can be used, to generate disruptions that occur either within the rime itself or between the onset and the rime. Thus, a TL manipulation of *train* can generate *trian* and *tarin*. If a difference in these two types of manipulation is shown, this is support for the notion that rime exerts an influence at a very early stage of visual word recognition.

Although there are many reports of TL effects in the literature, there are no reports of attempts to manipulate onsets and rimes in this way. We know that internal but not external transpositions prime (Chambers, 1979; Holmes & Ng, 1993; Perea & Lupker, 2003a, 2003b) and that TL words have effects similar to their base words (Andrews, 1996; Chambers, 1979; Forster, Davis, Schoknecht & Carter, 1987; Holmes & Ng, 1993; O'Connor & Forster, 1981; Perea & Lupker, 2003a, 2003b, 2004). While it is clear that lexical processing can tolerate a certain amount of letter position uncertainty, the question of interest here is whether that uncertainty is qualified by the onset-rime distinction; that is, are *trian* and *tarin* processed differently? If they are, then this would suggest that rime is an early constraint on word recognition; and the rime effects reported earlier are attributable to a prelexical stage in word recognition, rather than to the suggested feedback effects from orthography after the activation of lexical phonology.

Evidence that not all string-internal transpositions are equivalent was offered by Christianson et al. (2005) who showed that letter transpositions at internal morphological boundaries affected processing more than transpositions that did not cross such boundaries (for example, *susnhine* was more disruptive as a prime to naming than *sunhsine*). They showed that these effects extended to derivational morphemes (e.g.*float –er*) but not to pseudoderivation (e.g. *flound-er*), and argued that these findings indicate that string-internal letter positions are not all equal. However, these findings relate to morphological effects; the hypothesis tested in Experiment 6 is that letter position transpositions are also sensitive to onset-rime status, and as such, is a prediction for a prelexical effect.

Because of the nature of the task, Frankish and Turner (2007) argue that errors are a more appropriate dependent variable rather than latencies. If there is equivalent positional uncertainty for all letters between the first and last, then it should not matter where the TL manipulation occurs. If there is a subroutine that assigns letters to more or less likely onset and rime positions, then it would be expected that there would be a difference in errors depending on whether the rime alone or the onset and rime together have been disrupted. So the greater the disruption, the fewer the false positives; therefore, with errors as the DV, we would expect to see more errors with the within-rime manipulation, since it is arguably less disruptive.

Summary

TL manipulations enable an exploration to be made of the role of onset and rime in the very early stages of processing; differences in type of manipulation will strengthen the view that rime effects detected in Experiments 4 and 5 are attributable to this early processing stage rather than to feedback from orthography via lexical or sublexical phonology.

4.3.1 Method

Participants

35 University of Bristol students participated in return for partial course credit.

Apparatus

The equipment set-up, were the same as for the previous experiments; responses were made by means of a key press.

Stimuli and design

150 TL words were constructed from monosyllabic five- and six-letter words with a log frequency range of 0 – 888. TL manipulations were made at letter positions 2 - 3 (e.g. *dtich* and *durnk*) and 3 - 4 (e.g. *snkae* and *stirct*) so as to provide an equal number of rime-only and between onset and rime manipulations. 100 words with a similar frequency range were also used. Words and nonwords were pseudo-randomly ordered in five blocks of 50 items each. For stimulus lists, see Appendix E.

Procedure

Each trial began with a display of a fixation cross in the centre of the screen for 500 ms, followed by a blank screen for 200 ms. The target letter string was then displayed for 20 ms, and was immediately followed by a pattern mask for 250 ms. Letter strings were black lower case characters in 36 pt Times New Roman on a grey background. The pattern mask was made up of jumbled letter fragments, with an overall height and width 25% greater than the dimensions of the target display area. The task instructions were to classify each of the briefly displayed letter strings as word or nonword. There were no time limits for responding on each trial although

participants were advised to respond without delay on the basis of the impressions of what they had seen, rather than using complex problem solving strategies.
The initial block of 20 practice trials was followed by five blocks of 50 trials each. At the end of each block participants were encouraged to take a brief rest before proceeding to the next block.

4.3.2 Results

One stimulus was omitted from the analysis because of a TL typing error. The error rates for words and nonwords were 15% and 53%. Means for the errors for within-rime and between rime and onset nonwords are shown in Table 25.

Table 25

Percentage errors and standard deviations in lexical decision errors for TL words

	Mean errors	S.d.
TL within rime	52.5	15.8
TL between rime and onset	52.9	17.4

Clearly there was no difference in the mean number of false positives to the two different types of TL manipulation. In order to check that results were not confounded by stimulus length, position of TL change, base word frequency and bigram frequency, correlations were carried out but they were all non-significant. Pronounceability was established as an important factor in errors in Frankish and Turner (2007) so an analysis of responses by pronounceability was carried out. Table 26 shows that unpronounceable items elicited more errors than the pronounceable items: $t1\ (34) = 6.62, p < .001$, $t2\ (147) = 3.56, p < .001$. For the pronounceable items, there was no difference in errors to the different rime manipulations, but for the unpronounceable items, there were more errors where the TL manipulation included both the rime and the onset (significant by participants and marginal by items: $t1\ (34) = 4.2, p < .001$; $t2\ (42) = 1.9, p = .064$).

Table 26
Error rates for pronounceable and unpronounceable items

	Pronounceable Mean errors (%)	Unpronounceable Mean errors (%)
All items	50 (n = 105)	58 (n = 44)
TL within rime (e.g. *froce/hdege*)	48 (n = 42)	56 (n = 32)
TL between rime and onset (e.g. *balme/tiwst*)	50 (n = 63)	69 (n = 12)

4.3.3 Discussion

There was no effect of rime manipulation overall, so the hypothesis that the onset-rime framework might guide an early prelexical stage of letter string processing has to be rejected. There was a strong effect of pronounceability, with unpronounceable items generating more errors than pronounceable. Specifically, this finding replicates those of Frankish and Turner (2007) and, more generally, provides additional evidence for the activation of phonology at a very early stage in word recognition.

The TL manipulation aimed to uncover an onset-rime effect at this very early stage, on the grounds that first, the visual word recognition system might make use of an abstract onset-rime framework to help guide prelexical processing, and second, this mechanism could account for the rime effects seen in Experiments 4 and 5. There was some evidence for such a framework coming into play, but it was constrained by pronounceability. For pronounceable items, there was no effect of the rime/rime plus onset manipulation. For the unpronounceable items, however, the rime plus onset manipulation resulted in more false positives than the rime only manipulation. It was predicted that this condition would actually produce fewer errors than the rime-only condition, since it is arguably more disruptive and should therefore cause greater conflict with the base word and consequently fewer errors. However, the results were in the opposite direction; for the unpronounceable items, there were a significantly higher number of false positives for items where both the rime and onset were manipulated.

The prediction was based on the assumption that errors arise from matching with base words, but it is possible that the errors arise from processing that occurs before a lexical match is made. If all letter positions between the first and last letters have equal status, then it should not matter where adjacent letters are transposed (assuming, to take account of left-to-right processing, that letter position is controlled). But if the onset and rime structure acts as a framework or heuristic, then one might expect to see fewer errors where internal rime letters occupy any potential internal rime position, and more errors where the internal rime letters no longer occupy an appropriate rime-internal position. This is what the data show: so, for instance, *flase* generates fewer errors (16/35) than *tiwst* (30/35) because the l and a elements of *false* are still in potential rime letter positions; but the i of *tiwst* is no longer in a potential rime position for that rime. An alternative interpretation might be that the rime plus onset manipulation is more likely to produce strange letter combinations than the rime-only manipulation, but this seems unlikely since mean log bigram frequency was 10.04 for both conditions.

Caution needs to be exercised in interpreting these data, for several reasons. First, the analysis was carried out *post hoc*, including the pronounceability judgements. Second, the group of unpronounceable items was quite small and although the effect approached, it did not reach, significance by items. However, we might tentatively suggest that when phonology is not automatically activated for a letter string, a default mechanism comes into play in order to carry out orthographic parsing of a novel letter string. The unpronounceable onset-rime-manipulated items represent a very difficult problem for the visual word recognition to solve; however, this extreme problem is not typical of the type of items that have been the subject of the experiments so far, particularly in terms of pronounceability. For pronounceable items, the rime manipulation had no effect, so on the basis of this experiment we can conclude that the rime effect seen in Experiments 3 and 4 did indeed arise from sublexical or lexical orthography, after the activation of phonology, and not from a prelexical parsing process.

4.4 Summary

This chapter explored the role of rime; Experiment 6 showed that readers are sensitive to the probability of rime spelling, an effect that emerges from lexical or sublexical processing. Experiment 7 explored whether the rime effect could emerge from prelexical processing, but there was little evidence for this. Using the masked lexical decision technique with TL words as stimuli provided more evidence for automatic activation of phonology, even at very short presentation times. Overall, there is evidence that the rime spelling is a powerful source of information for readers and that where the rime spelling is unusual, processing is slowed. Experimental nonword stimuli are often characterised by strange rimes; the evidence from these experiments suggests that responses to such spellings are slower than normal.

Chapter 5

Masked priming with wordlike stimuli

"The heart of our trouble is with our foolish alphabet.
It doesn't know how to spell, and can't be taught."
(Mark Twain).

The previous chapter investigated the role of rimes; it was shown that readers are sensitive to rime spelling in lexical decision, and Experiment 6 investigated the possibility that onset-rime framework is of importance in the very early stages of visual word processing. The conclusion from that experiment was that a rime effect was not apparent at this early stage and that generally the onset-rime structure does not guide graphemic parsing; although there was a suggestion that unpronounceable letter strings might activate some attempt to parse using an onset-rime structure. An alternative approach that also prevents participants from becoming alert to the nature of the experimental manipulation is to use the masked priming technique. Again, this enables the very early stages of word processing to be investigated, and, in the case of rimes, has the additional advantage of keeping the orthographic input intact rather than changing it. The focus of the first part of this chapter is on two experiments using the forward masked priming technique.

5.1 Masked priming effects – what are they and why do they matter?

Masked priming involves presenting participants with a prime that is masked and presented so briefly that it is unavailable for conscious report, but nevertheless affects responses to a target item that is some way related to the prime. While the notion that this procedure uncovers mechanisms that operate at the earliest stage of visual word recognition is uncontroversial, it is matter of debate as to what the effects actually are. For example, Coltheart et al. (2001), when discussing whether the DRC model could simulate masked priming effects, commented that there were difficulties "concerning exactly what the effects are that would need to be simulated" (p. 250). In a similar vein, Frost, Ahissar, Gotesman, and Tayeb (2003, p. 48) noted that the literature is characterized by inconsistent results. The key issue is whether priming effects are primarily orthographic (as argued by, for example, Bodner & Masson, 1997) or whether phonology also exerts an effect, even at this very early stage (e.g. Ferrand & Grainger, 1992; Perfetti, Bell & Delaney, 1988).

Findings have been difficult to interpret because results appear to depend on a number of procedural factors, such as the lighting conditions in the testing room (Lukatela, Frost & Turvey, 1998, 1999) or stimulus luminance (Tzur & Frost, 2007). In a meta-analytic review, Rastle and Brysbaert (2006) concluded that, nevertheless, a small but statistically reliable phonological effect of about 10 ms occurs in forward masked primed lexical decision (based on studies by Berent, 1997; Bowers, Vigliocco & Haan, 1998; Davis, Castles & Iakovidis, 1998, Expt. 1; Grainger & Ferrand, 1994; Holyk & Pexman, 2004; Lukatela et al., 1998; Lukatela & Turvey, 2000). Those earlier experiments compared lexical decisions to words when primed with either pseudohomophones or graphemic controls (e.g. *cake* – *kake/pake*), and found that responses were faster to words primed by pseudohomophones than to words primed by graphemic controls. Rastle and Brysbaert carried out two new experiments and reported phonological priming effects, which they interpreted in terms of a compromise between the 'strong' and 'weak' views of phonology. The 'strong' view is that priming emerges from a fast phonological system, while the standard 'weak' explanation is that the apparent phonological priming effect emerges in the orthographic lexicon via feedback from phonology. After a series of unsuccessful simulations with the DRC model, the authors concluded that the effect emerges as a result of the level of activation in the phonological lexicon, but that this information is constrained by orthographic information. The information in the strength of the signal is sufficient to give correct decisions; for example, both *brain* and *brane* generate /breɪn/ in the phonological lexicon, but *brain* generates a stronger signal because it generates activation via both pathways, while *brane* only generates activation via the assembled route. Simulations with the DRC model and an investigation of activity in the phonological lexicon supported this interpretation.

However, many of the pseudohomophone and nonword stimuli used by Rastle and

Brysbaert were not wordlike, and it is arguable that the results they reported were attributable to a factor other than lexical homophony. There have been several demonstrations in the previous chapters that lexical phonology is less likely to be activated by unwordlike than wordlike stimuli. For example, the weaker readers in Experiment 4 were more likely to incorrectly reject unwordlike pseudohomophones in phonological lexical decision, suggesting that these items did not easily activate phonological representations; and the analysis of the wordlikeness ratings showed that there was no base word frequency effect for the 100 unwordlike items, although there was an effect for the 100 most wordlike.

While it is generally informative to establish whether a priming effect can be achieved using more wordlike stimuli, the more specific question of interest here is to determine whether an orthographic rime effect can be detected in this task. If the rime effect is prelexical, that is, if the system is sensitive at an early stage to more or less frequently encountered spelling patterns, then we can predict that dominant rimes will speed responses to subdominant targets, irrespective of lexical status; words, nonwords and pseudohomophones should all show similar effects. If the rime effect, detected as an important component in responses in Experiment 4, arises from sublexical orthographic feedback once a stable phonological representation has been activated, then responses will differ according to lexical status. Rime effects will be apparent for words but not for nonwords. Predictions regarding pseudohomophones are less clear; most research has not investigated what happens when pseudohomophones are primed, probably because of the influential view, arising from the dual-route approach, that only words prime. However, Lukatela, Eaton, Lee, Carello and Turvey (2002) established that pseudohomophones negatively prime pseudohomophones in lexical decision, so it seems likely that word primes should exert a similar effect. Words with dominant orthographic rimes are likely to exert a stronger effect on subdominant-rime pseudohomophones than words with subdominant rimes priming dominant-rime targets.

5.2 Experiment 7

Masked lexical decision with dominant and subdominant primes

Summary and predictions

Masked priming is a useful technique to investigate unconscious processes in visual word recognition. Homophonic primes have been shown to affect lexical decision and therefore offer support for the view that phonology is automatically activated with the possible concomitant activation of base word orthography. This experiment aims to use wordlike stimuli in a primed lexical decision experiment, with the broad prediction that homophonic primes will affect response times and errors to words; and more specifically, that orthographic rime may also affect responses.

5.2.1 Method

Participants

32 University of Bristol students participated in return for partial course credit.

Apparatus

The equipment set-up was the same as for the previous experiments.

Stimuli and design

50 monosyllabic words with a $\log_{10}$ frequency range of 0 – 2.69 were chosen as the target stimuli, 25 each with dominant/subdominant rimes (mean frequency 1.03, 1.1). Pseudohomophone primes were created by changing the rime to its primary alternative (legitimate) spelling. 50 monosyllabic pseudohomophone targets were created from base words with the same frequency range, also with an equal mix of dominant and subdominant rimes (mean frequency 1.14, 1.15). Primes for these items were the base words. Unpronounceable orthographic controls were created for both types of target, controlling for the number of shared letters (position-sensitive) between primes and targets. The list therefore contained 8 types of item, as in Table 27. (For complete list, see Appendix F.)

Table 27
Stimuli for Experiment 7, masked primed lexical decision.

Prime	TARGET	Example
Subdominant PH	Word - Dominant rime	bleek - BLEAK
Graphemic control	Word - Dominant rime	blemk - BLEAK
Dominant PH	Word - Subdominant rime	chare - CHAIR
Graphemic control	Word - Subdominant rime	chaje - CHAIR
Subdominant word	PH - Dominant rime	boast - BOST
Graphemic control	PH - Dominant rime	boavt - BOST
Dominant word	PH - Subdominant rime	churn - CHERN
Graphemic control	PH – Subdominant rime	chsrn - CHERN

PH = pseudohomophone

The final list of 200 items was presented to all participants, which meant that they saw all stimuli twice, once with a homophonic prime and once with an unpronounceable graphemic control prime. The advantage of this design is that it eliminates between participants' variation, and repetition of stimuli within long lists is not considered to present a problem in this type of experiment (e.g. see Bowers, Vigliocco & Haan, 1998, for a precedent). Six randomly ordered lists were created with the constraints that there were no sequences of more than 3 items eliciting the same response, and presentations of the same target were at least 60 items apart. Stimulus presentation and data recording were accomplished via the DMDX software (Forster & Forster, 2003).

Procedure

While there are several alternative experimental priming paradigms, the forward masked technique was chosen for this and the following experiment, mainly because it is the same procedure as that chosen by Rastle and Brysbaert and therefore presents a useful comparison. In forward masked priming, a pattern mask is followed immediately by a briefly presented prime, and this is in turn masked by a target presented in different case.

Participants were tested individually in a dimly-lit room. They were told that they would see a series of letter strings in upper case and that they would be required to decide whether each one was a word as quickly and as accurately as possible.

Participants were told that each letter string would be preceded by a series of hash marks, but they were not told of the existence of the prime. All primes were presented in lower case for 50 ms and preceded by a mask of hash marks lasting 500 ms, and was followed immediately by a target in upper case, also presented for 500 ms. Participants were presented in rotation with one of the six lists; they carried out 12 practice trials before the main experiment.

5.2.2 Results

Errors were trimmed by 3.00 standard deviations. Trimming and incorrect responses accounted for a loss of 19.5% of the total data. Items with an error rate of more than 50% were removed from the analysis (words *deem, goad, maim, strait, waif* and *bream*; and pseudohomophones *cheak, screach, snear,* and *sweed*). Data from seven participants who, after these items had been removed, still had a high error rate (> 20%) were also removed, leaving 25 participants' data for analysis. Finally, items were individually trimmed where they showed a mean reaction time of more than 1200 ms. 15.2% of the results in the final data set were classed as errors.
Table 28 shows that words were responded to faster than pseudohomophones. This difference was significant: t1 (99) = 10.53, $p < .001$; t2 (178) = 12.58, $p < .001$.

Table 28
Mean decision times (ms), standard deviations and errors in the primed lexical decision task.

	Mean RT (ms)	S.d.	Errors (%)
Words	662	131	12.7
Pseudohomophones	756	138	17.4

Prime type (homophonic or orthographic control) did not affect responses to words (see Figure 19: all ps >.11 by participants, >.9 by items).

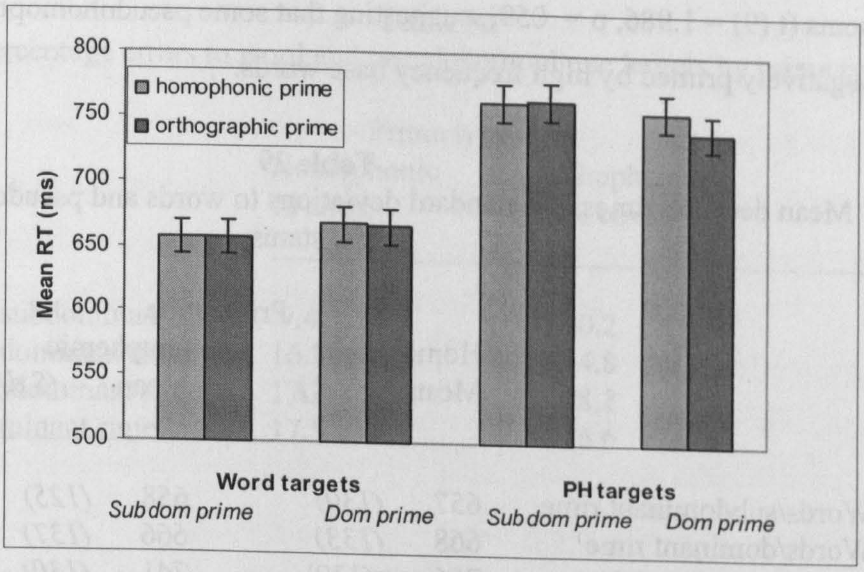


Figure 18. Responses to words and pseudohomophones with dominant and subdominant rimes: mean reaction times and standard error of the means.

Irrespective of priming, there was a tendency for rime-dominant items to be responded to more slowly. Table 29 shows that decisions to words with a dominant rime were on average 10 ms slower than words with a subdominant rime, but this difference was not significant by items, although it was marginal by participants: F1 (1,24) = 4.18, p = .052; F2 (1,86) = .023, n.s. Pseudohomophones with dominant rimes also elicited slower responses than those with subdominant rimes; the difference was significant by participants but not by items: F1 (1,24) = 6.114, p = .021; F2 (1,90) = .629, n.s. *Post hoc* tests showed differences between the two graphemic control conditions that were also significant by participants but not by items: t1 (24) = 2.183, p = .021; t2 (44) = 1.583, n.s.)

Responses to pseudohomophones with subdominant rimes were slowed by 15 ms when primed with a dominant-rime word compared to the orthographic control condition; this difference was significant by participants but not by items (t (24) = -3.183, p = .004; t (45) = .98, n.s.). For the 10 highest-frequency of these items, the difference was significant by participants (t (24) = 7.975, p = .01) and, although not significant by items, this was probably attributable to one outlier (*felt*/FEALT) which showed a 108 ms difference in the opposite direction to the nine other high frequency items. When this item was removed, the difference was marginally significant by items (t (9) = 1.986, p = .059), suggesting that some pseudohomophones may be negatively primed by high frequency base words.

Table 29
Mean decision times and standard deviations to words and pseudohomophones by rime status.

	Prime type				
	Homophonic		Graphemic		
	Mean	*(S.d.)*	Mean	*(S.d.)*	Diff
Words/subdominant rime	657	*(130)*	658	*(125)*	-1
Words/dominant rime	668	*(133)*	666	*(137)*	2
PHs/subdominant rime	756	*(139)*	741	*(130)*	15
PHs/dominant rime	762	*(143)*	763	*(140)*	-1

Reader speed effects were shown in the results of Experiment 3, so it is possible that significant effects are masked by individual differences in participants, so latencies were re-analysed by reader speed. While there were no effects for words or pseudohomophones for the fastest readers, the ten slowest readers did show an effect for the pseudohomophone/subdominant rime condition that was marginal by participants, t (9) = 1.995 p = .077 and significant by items, t (24) = 2.65, p = .014. This indicated that the presentation of a prime with a dominant rime tended to delay 'No' responses to the subdominant target pseudohomophone. There were no effects for the slower participants in the other pseudohomophone condition, or in either word condition.

Error analysis

The error rate was high, but there was no clear pattern of effects (Table 30). There were no significant differences between conditions. Since errors in Experiment 3 had shown correlations with base word frequency, similar correlations were carried out here. There were significant correlations between frequency and word responses, but the relationships between base word frequency and pseudohomophone responses were much weaker (Table 31).

Table 30
Percentage errors to word and pseudohomophone targets by prime type.

	Prime type Homophonic % errors	Graphemic % errors
Words – subdominant rime	9.4	10.2
Words – dominant rime	16.3	14.8
PHs – subdominant rime	17.3	18.2
PHs – dominant rime	17.5	16.6

Table 31
Correlation between errors and base word frequency by prime type.

	Prime type Homophonic r value	Graphemic r value
Words – subdominant rime	-.64**	-.50**
Words – dominant rime	-.64**	-.70**
PHs – subdominant rime	-.35 p <.1	-.35, p <.1
PHs – dominant rime	-.19, n.s.	-.12, n.s.

** = p <.05, ** = p <.001*

5.2.3 Discussion

The immediate aim of this experiment was to investigate whether prime rime status affected responses to targets in lexical decision; a clear effect would have been evidence for a very early influence of rime, and a null effect would offer support for the view expressed in Experiment 5 that rime effects emerged from bi-directional links from lexical phonology to base word orthography. This experiment has not shown an influence of rime overall and, therefore, taken together with the results of Experiment 6, this suggests that rime effects do not occur at this early stage in the visual word recognition system. However, results may have been affected by individual differences; there was a wide variation in responses (as indicated by the standard deviations in Table 29) among the final 25 participants. Indeed, it must be suggested that, given the small effect sizes reported by Rastle & Brysbaert (2006) in masked phonological priming experiments, there is not enough power in this experiment to uncover a priming effect. Nevertheless, it is possible that there was an effect of rime dominance since responses were slowed for pseudohomophones with subdominant rimes when they were primed with rime-dominant words, and this was close to significance for slower participants, and also for the higher-frequency items. There was also a weak correlation with base word frequency for the subdominant pseudohomophones, but not for the dominant items, indicating that the base word may have been exerting some effect on responses.

A cautious interpretation of these findings suggests that words with strong links (i.e. those with dominant rimes) generate the strongest possible 'yes' signal as primes, and this signal conflicts with that from the target and so correct 'no' responses are slowed. This effect is strongest for high frequency items – the combination of frequency and rime dominance means that the spelling for these items is very secure. Priming in this case is inhibitory; it presumably stems from lexical activation rather than sublexical (or a similar result would be shown for the nonword and pseudohomophone primes). It is difficult to conclude that it is phonological, however, because the delay in pseudohomophone responses could simply arise from the conflict with the prime spelling; and it is possible that the effect is orthographic, and similar to results reported by Segui and Grainger (1990), who established inhibitory effects for high frequency neighbours on low frequency targets. This effect was most noticeable among the slowest readers; presumably because, as we have already seen, this group is less able to suppress sources of conflicting information (see Experiment 3, Section 3.2.2). There was no effect for pseudohomophones with dominant rimes, because words with subdominant rimes have less strong spelling patterns, and do not signal 'yes' so strongly; therefore they do not conflict as strongly with the 'no' response to their targets. However, these results, and their interpretation, must be treated with caution, because there were no results that were significant both by participants and by items.

Little evidence for a priming effect was shown in this experiment. One possibility is that this may be attributable to the nature of the primes. In this experiment, the control primes were difficult to pronounce, or unpronounceable, but in earlier research, including the experiments reported by Rastle and Brysbaert (2006), orthographic controls were pronounceable. If pronounceable control primes activate a phonological representation that conflicts with the target phonology, then this must add to the processing time. This negative priming could not have occurred in this experiment, since the primes were unpronounceable, and therefore it is possible that the baseline control responses times were set higher here than in previous work, and this of itself accounts for the failure to detect homophonic priming.

Nevertheless, one of the major lines of argument in this work has been that wordlike items are more likely to activate phonology than unwordlike stimuli, and so one might still have expected to see an effect of homophonic priming. The absence of evidence in this experiment therefore presents a something of a challenge to this argument. One explanation that might account for the null effects can be couched in terms of the experimental design; this was such that all targets had phonological lexical status. A similar design was used by Rastle and Brysbaert (2006) for their second experiment, and they noted that "phonological recoding in this situation makes the word-nonword discrimination particularly difficult" (p. 117). Decisions based on phonology alone would all generate 'yes' responses, and since this would be incorrect for the pseudohomophone targets, a strategy that involved suppressing sources of phonological information, while prioritising orthographic information, would be most helpful in making correct decisions. This strategy might extend to suppressing any information from the prime as well as the target. Evidence for a residual effect of homophonic priming came from the high-frequency, rime-dominant words priming subdominant-rime pseudohomophones, and it is precisely this set of primes that might be most resistant to phonological suppression.

An explanation couched in terms of the design can also account for the high error rate and slow latencies. Response times were certainly longer than those reported in previous research (as summarised in Rastle & Brysbaert, 2006), and the error rate (15.2%) was also high by comparison with previous work. This was clearly a hard task for participants to do; students from Bristol University would be expected to have good vocabulary knowledge, but seven had to be excluded from the final analysis because they made errors on more than one in five items. The problem seems to be located in the wordlikeness of the target stimuli; if all the targets, irrespective of phonology, look more or less like words, then lexical decision must be harder than when half the items look strange. Support for this view comes from the study reported by Vanhoy and Van Orden (2001) who found that items with existing rimes, such as *jale*, elicited longer latencies and more errors in lexical decision than items with non-existing rimes, such as *jael*. Their interpretation was that *jale* activated a more coherent resonance with the base word than *jael*. Since all the target items in this experiment were analogous to *jale*, in having existing rimes, this would have led to longer response times and more errors. Suppressing sources of phonological information would have exacerbated the difficulty of dealing with these items. Thus, even though phonological activation is theoretically unhelpful in visual lexical decision, it is necessary for accurate pseudohomophone rejections; otherwise all decisions have to be based on orthography, and where all the stimuli look like words, these decisions become very difficult to make. When spelling is plausible, phonological activation is therefore not only mandatory and automatic, it is essential.

Summary

In summary, Experiment 7, as Experiment 6, gave no indication that the orthographic rime exerts a very early, prelexical, effect in lexical decision. There was also no overall effect of homophonic priming; but the absence of clearly detectable effects was possibly attributable to the design of the experiment, or to inadequate power. The wordlikeness of the stimuli made decisions based on orthography very difficult, leading to slow response times and a high error rate.

5.3 Experiment 8

Masked priming with wordlike and unwordlike primes

The final experiment brings together the two main threads that have been running through the previous experiments; first, that wordlike items are processed differently from unwordlike, and second, that the rime orthography is a powerful predictor of wordlikeness. We have seen some evidence for the importance of rime in both the experimental results and in the wordlikeness ratings. The final experiment explores whether wordlike pseudohomophones are more effective primes than unwordlike. If wordlike pseudohomophones are more likely to activate base word phonology than unwordlike, we would expect items like *brane* would be more effective primes than items like *brayn*. However, several researchers have reported masked phonological priming using lists containing unwordlike items; for example, Rastle and Brysbaert (2006) reported significant phonological priming effects from lists containing items such as *wreighed* and *phoak*. If such stimuli can be shown to exert a priming effect, then more wordlike items such as *rade* and *foke* should demonstrate an even stronger effect. Therefore, the aim of the final experiment was to establish whether wordlike and unwordlike primes gave different priming effects. If wordlike items are shown to be more effective primes than unwordlike, this is further support for the argument that stimuli used in psycholinguistic experiments need to be orthotactically plausible in order to reveal the workings of visual word recognition mechanisms.

An immediate problem is that Experiment 7 did not give evidence of a phonological priming effect. The reason for this may be attributable to the design; and/or because the prime presentation time of 50 ms did not give time for a complete phonological representation of the prime to be established. A number of authors have suggested that phonological effects occur later than orthographic; for example, Grainger, Diependaele, Spinelli, Ferrand and Farioli (2003) suggested that there were no significant pseudohomophone effects for prime durations of less than 60 ms in lexical decision tasks, a finding supported by Lee, Kambe, Pollatsek and Rayner (2005) who found that there was no priming for pseudohomophones in reading and naming tasks at durations under 60 ms. Similarly, Lukatela and Turvey (1994b) found a pseudohomophone priming effect at 250 ms but not at 60 ms. Reporting an event-related potential study, Grainger, Kiyonaga and Holcomb (2006), suggested that orthographic processing effects arise around 33 ms, while phonological effects are not fully-established until around 67 ms. There is certainly evidence for an early stage of orthographic processing that is separate from phonological: so that, for example, *grdn* primes GARDEN and *blcn* primes BALCON (Grainger, Granier, Farioli, Van Assche & van Heuven, 2006; Peressotti & Grainger, 1999); and Bodner and Masson (1997) suggested that the main effect of primes is to facilitate orthographic encoding, so priming is essentially nonlexical (p. 268). Standing as counterparts to Rastle and Brysbaert's findings, such studies suggest two possible alternative interpretations of masked priming effects; the first is that phonological priming occurs later than orthographic and will only be apparent with longer prime durations; *the second is that phonological priming does not occur at all*, which implies that the effects that have been reported in the literature may be attributable to orthographic factors (e.g. Frost et al. 2005, who found robust orthographic priming effects in English by comparison with Hebrew).

If pseudohomophone priming does exert an effect on target words, but is dependent on prime duration, then we would expect to see more priming for wordlike items than unwordlike at short presentation times, because wordlike items are better at activating phonology. On the other hand, if what is apparently phonological lexical priming is actually orthographic, then we would still expect to see wordlike primes priming better than unwordlike, because they do not contain the unusual letter sequences apparent in non-existent rimes (e.g. final *-j*) which were markers of wordlikeness ratings. However, if wordlikeness effects are not detectable at this very early stage of visual word recognition, then there will be no difference between wordlike and unwordlike items, and we must conclude that readers' sensitivity to orthotactic regularities is a feature of visual word recognition that occurs at a later stage.

Summary and predictions

If the visual word recognition system is sensitive to orthotactic regularities in nonword stimuli at a very early stage, then wordlikeness effects should be detectable in a masked priming experiment, such that wordlike pseudohomophones should be more effective primes of target words than unwordlike stimuli. If the primes are constructed so as to be equivalent in their orthographic relation to the target words, then any priming advantage must be phonological rather than orthographic.

5.3.1 Method

Participants

51 University of Bristol students participated in return for partial course credit. Upon arrival at the lab, they were randomly assigned to one of the experimental conditions or to the control condition, until there were 18 participants in each experimental condition, and 15 in the control group.

Apparatus

The equipment set-up was the same as for the previous experiments.

Stimuli and design

The target stimuli were 80 monosyllabic words with a $\log_{10}$ frequency range of 0.3 – 3.08; pairs of wordlike and unwordlike primes were created for each word (e.g. *spunge, spunj* - SPONGE). The 160 primes were rated for wordlikeness by 20 students and colleagues of the researcher, and ratings calculated as described in Section 2.2.1. In order to overcome the problem identified in Experiment 7, of decreased reliance on phonology, the control targets were true nonwords rather than pseudohomophones; lexical decisions would therefore be facilitated by phonological activation. 80 nonwords were chosen from the stimuli used in Experiments 1 and 2 so as to have a similar wordlikeness range as the new items, and homophonic primes for the nonwords were created by changing one letter (e.g. *barce* - BARSE). (See Appendix G for stimuli).

Two experimental lists containing the 80 target words and 80 nonwords were created. The target words were primed by either the wordlike or the unwordlike pseudohomophone of the pair, with an equal number of each pair in each list. A control list of the same 80 words, primed by pronounceable but non-homophonic nonwords, plus the 80 nonwords, was also created. Targets were randomly mixed, as before, with the nonwords; each participant was presented with only one of the three lists. Stimulus presentation and data recording were accomplished via the DMDX software (Forster & Forster, 2003).

Procedure

Participants were tested individually in a dimly-lit room. They were told that they would see a series of letter strings in upper case and that they would be required to decide whether each one was a word as quickly and as accurately as possible. Participants were told that each letter string would be preceded by a series of hash marks, but they were not told of the existence of the prime. Each prime was presented in lower case for 50 ms and preceded by a mask of hash marks lasting 500 ms, and was followed immediately by the target in upper case, also presented for 500 ms. Participants were presented with 10 practice trials before the main experiment.

After the main experiment, participants were debriefed and then shown the primes, which they were asked to rate according to visual wordlikeness on a 1 – 7 scale. Their ratings were then converted into wordlikeness z scores, using the procedure described in Section 2.2.1. Correlations between these ratings and those gathered as part of the stimulus construction process were all >.88. The participants' ratings were those used as the wordlikeness measure in the results analysis.

5.3.2 Results

Errors were trimmed by 3 standard deviations. Any response with a mean reaction time of more than 1000 ms was also removed. No individual stimuli were removed because there were no clear patterns of errors; 9.2% of the items in the final data set were classified as errors.

Table 32
Mean decision times (ms), standard deviations and percentage errors for words and nonwords by list type.

	Words Mean RT *(S.d.)*	% errors	Nonwords Mean RT *(S.d.)*	% errors	RT Diff
List 1	566 *(54)*	9.9	621 *(39)*	8.5	-45
List 2	538 *(42)*	9.7	586 *(37)*	8.6	-48
List 3 (control)	591 *(66)*	9.4	623 *(46)*	10.5	-32
Grand mean	565 *(54)*	9.7	610 *(40)*	9.2	-45

Table 32 shows that words in all conditions were faster than nonwords; on average, by 45 ms. An analysis of variance, in which target type was treated as a repeated factor and list version (3 levels) as an unrepeated factor was significant by participants and by items: $F1\ (1, 48) = 97.7$, $p < .001$: $F2\ (1, 237) = 100.9$, $p < .001$. In addition, *post hoc* Bonferroni tests indicated that there were significant differences between all three lists for the by items analysis: all ps < .037). The fact that there was a significant difference in responses to the two experimental lists (1 and 2) is a potentially confounding factor in the analysis, so latency data in all conditions were recoded as z scores in order to place all reaction times on the same scale.

There were no significant differences in errors to words and nonwords: F1 (1,48) = .32, n.s.; F2 (1,237) = .31, n.s.. Overall, the error rate was much lower than that in Experiment 7, so it appears that the design, which was created so as to prioritise phonological processing and thus facilitate accurate judgements, was successful.

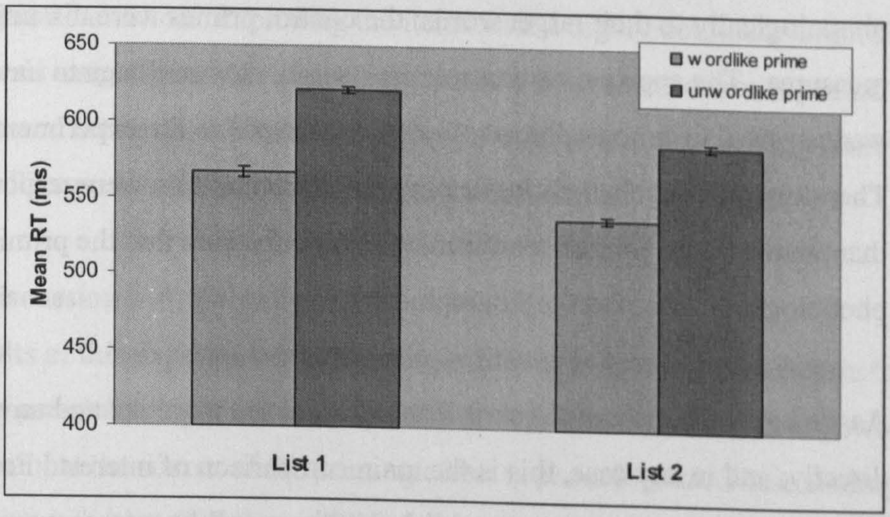


Figure 19. Decision times (ms) to words primed with wordlike and unwordlike pseudohomophones.

Figure 19 indicates that wordlike pseudohomophones facilitated decisions over unwordlike primes in both experimental lists; planned t-tests comparing z score reaction times for all responses were significant by participants but not by items: t1 (35) = 4.4, p < .001; t2 (158) = .66, n.s.. Given a strong by participants effect, the by items analysis should be considered further. It is likely that the driving force behind decisions to target words is word frequency, and if responses are primarily driven by word frequency, then this may mask any secondary priming effect; for example, responses to high frequency target FIRST will always be faster than responses to low frequency target WEEP, regardless of prime type. And indeed, correlations show that reaction times were faster for higher frequency targets for both types of prime: wordlike, r (80) = -.34, p = .002; unwordlike, r (80) = -.35, p = .002. Therefore a repeated measures test, controlling for base word frequency, is the more appropriate statistical test; and this analysis revealed a significant effect of wordlikeness: $F2\ (1, 78) = 10.19$, $p = .002$.

Orthographic issues

An immediate problem with analysing the results of this experiment is that it was not possible to make a direct comparison with the control group's results, since the primes in the experimental and control conditions were not orthographically equivalent. The pseudohomophone primes were similar orthographically and phonologically to their target words; the control primes were dissimilar on both measures. The appropriate control measure should have been to have primes that were as similar orthographically to the target word as the experimental primes. Therefore, although words in the experimental condition were responded to faster than words in the control condition, we cannot be sure that the priming effect was phonological rather than orthographic.

As we have seen, it is still possible to compare the wordlike and unwordlike primes directly; and in any case, this is the main comparison of interest. For this comparison to be valid, it is important to establish that the wordlike primes were not orthographically more similar to their base words than the unwordlike primes. If there was a difference, then it would have to be acknowledged that any difference in priming effects between the two sets could be orthographic rather than phonological. To determine this, a calculation was made of the "orthographic similarity" measure (OS) devised by Van Orden (1987) (see Appendix H for details of the calculation). This showed that both sets of primes were equivalent in their similarity to their base words (wordlike, .66, unwordlike, .64; where 0 represents no relationship and 1 represents a perfect similarity between two letter strings). If there is no difference in the orthography of the nonword primes as they relate to their target word, any difference between primes cannot be attributable to confounding differences in orthography. It seems likely that the speeded responses to the targets linked to the wordlike set must be attributable to the activation of base word phonology.

Rime analysis

The effect is, however, modulated by the effect of target word frequency. One might argue that this is because strongly-mapped spellings are more immune to priming effects than less strongly-mapped spellings. It is likely that the rime is the major source of this effect, given the evidence already provided in earlier chapters, together with the results for dominant-prime words seen in Experiment 7. Looking at the difference between the wordlike and unwordlike primes, there is a significant negative correlation with target rime dominance ($r(80) = -.34$, $p = .002$), indicating that the more dominant the target rime, the smaller the difference between the two prime types. Thus, if the target rime spelling is very strongly mapped, the phonological impact of primes (with, by definition, lower-frequency spelling patterns) is diminished. Excluding the items with a highly dominant target rime ($n = 19$), reveals a significant difference between response times to words primed with wordlike and unwordlike primes ($t2(61) = 2.66$, $p = .01$); once again, clear evidence that wordlike primes facilitated priming over unwordlike.

5.3.3 Discussion

The results of this experiment offer strong evidence in favour of the notion that wordlike pseudohomophones are more effective primes for words than unwordlike items, and that the source of the priming effect is phonological. These findings are rather more coherent than those of Experiment 7; this is possibly attributable to the use of wordlike nonwords as filler items, which meant that the best strategy to achieve correct answers would be to make use of phonological information. The error rate in this experiment was considerably lower than in Experiment 7, so this approach would appear to have been successful. Although much previous research would suggest that phonological priming would not be apparent at a prime duration of 50ms (e.g. Grainger et al., 2006; Lee et al., 2005; Lukatela & Turvey, 1994) the results show that wordlike primes exerted an effect; this was, however, modulated by target word frequency and rime spelling. Irrespective of prime status, there was a strong frequency effect in decisions to words, and this masked any priming effect; controlling for frequency revealed a difference between wordlike and unwordlike primes. Higher frequency words are associated with more strongly mapped spellings, so it is likely that these items are more resistant to priming effects. This was demonstrated by excluding those items with a highly dominant rime orthography; the remaining words clearly showed a wordlikeness priming effect. Thus, it appears that strongly-mapped spellings generated responses that were relatively impervious to the effect of primes; but in cases where there was likely to be more uncertainty about a word's spelling, responses were affected by the wordlikeness of the phonological prime.

Phonological priming effects have been elusive in previous research (Rastle & Brysbaert, 2006), and various factors have been suggested to account for the fragility of the effect, including: relatedness proportion (Brysbaert & Praet, 1992; Verstaen, Humphreys, Olson, & d'Ydewalle, 1995) , the lighting conditions in the testing room (Lukatela et al., 1998; Lukatela et al., 1999), and stimulus luminance (Tzur & Frost, 2006), and, explaining contradictory results from two virtually identical experiments, individual differences in phonological and perceptual skill (Holyk & Pexman, 2004). The results of this experiment suggest that there could be three more reasons for the unclear results. First, if words with very strongly-mapped rime spellings are used as targets, the effect will not be detectable. Second, if unwordlike primes are used, phonological activation may develop only slowly. Third, if unwordlike nonword filler items are used, this will lead to a general de-emphasis on phonological processing since decisions could effectively be made on orthography. Although it is somewhat depressing to add to the long list of factors that affect masked priming, these three new factors have the advantage of being theoretically embedded within the visual word recognition system itself (unlike external factors such as lighting and luminance).

The thorny problem of appropriate graphemic controls

In this experiment, control primes were orthographically very dissimilar from the target primes and therefore the comparison that is usually made in this type of experiment could not be carried out, since it would have been impossible to state whether any difference between the experimental and control conditions was attributable to orthographic or phonological priming. However, the comparison between wordlike and unwordlike items still enabled a clear demonstration of a phonological priming effect. As it happens, this may ultimately be a better test than a comparison with graphemic controls, because of the difficulties in establishing what constitutes an adequate control. Rastle and Brysbaert (2006) called this a "theoretical void" (p. 112) and cited a variety of different approaches based on different coding schemes (e.g. Bowers, 2002; Coltheart et al., 2001; Davis, 1999; Grainger & Jacobs, 1996; Harm & Seidenberg, 2004). In their experiments, Rastle and Brysbaert used a left-aligned slot-based coding scheme (such as that used by Berent, 1997; Davis et al., 1998; Perfetti et al., 1988; Perfetti & Bell, 1991; Rayner, Sereno, Lesch, & Pollatsek, 1995) such that graphemic controls preserved position-specific shared letters across primes and targets. Phonological priming effects would then reflect savings on target processing due to phonological overlap between prime and target once the savings from the orthographic overlap between prime and target had been eliminated. One issue with this approach is that it omits the effect of any phonological activation from the graphemic control itself. The assumption is that these controls do not activate phonology at all i.e. that while *cete* activates lexical phonology, its control *dest* remains purely orthographic. Given that *dest* is rather more wordlike than *cete*, it seems more than likely that *dest* would generate some phonological activation; which may then have an inhibitory effect on the target *seat*, and so slow responses. An additional problem is that this approach means that at least 10% of the graphemic controls are also pseudohomophones (e.g. *berne, broak, wreach, wribbed, wraik, thorde, toid, toye, heale, larc*) and these might also inhibit target responses. Therefore, what have previously been reported as phonological priming effects may in part at least arise from inhibitory activation from the control primes. The only way to deal with this confound is to use unpronounceable primes, that are as orthographically as similar to the target word as the prime (as in Frankish & Barnes, 2008), or possibly to use identity primes.

What do wordlikeness effects tell us?

This experiment has shown a clear effect of wordlikeness in that wordlike primes were more likely to show phonological priming than unwordlike. Previous failures to find phonological priming effects may be attributable to the use of unwordlike primes, which are characterized by unfamiliar orthotactic patterns. Wordlike pseudohomophones are constructed from commonly-encountered patterns and the indications are that these patterns activate phonological representations more quickly and more efficiently than unusual letter sequences. In broad terms, the explanation for phonological priming given by Rastle and Brysbaert can deal with these results; that is, priming effects emerge from activation in the phonological lexicon, activation which is constrained by orthographic information. This is acceptable in theory but in practice it does not capture the difference between wordlike and unwordlike primes, even though the difference should be contained within 'orthographic information'. Thus, as far as the DRC computational model is concerned, any difference between *spunj* and *spunge* will be such that *spunge* will be processed more slowly by the grapheme-to-phoneme route, because the final *–ge* involves a 'whammy'; and indeed a quick simulation with the DRC model produces *spunj* in 80 cycles, and *spunge* in 151. However, the evidence from human readers has demonstrated that not only is *spunj* processed more slowly (e.g. Experiments 1 and 3), it will be less likely to activate base word phonology (Experiment 8).

These findings offer evidence that letter-by-letter processing is, even at this very early stage, quickly superseded by graphemic processing. This is in opposition to the DRC's characterization of nonword reading processes, and other approaches that assume a primary role for letter-by-letter processing. (e.g. Pelli, Farrell & Niirem 2003). If in fact certain combinations of letters are encountered more often in one's reading experience than certain individual letters, then it seems more than likely that the system discovers that frequently-encountered patterns are more efficiently processed as integrated units rather than as separate letters. This strategy will be particularly useful in English since so many pronunciations have to be captured by so few letters; once letters start to form units, the task becomes more transparent. Some of this knowledge, that has developed through exposure to the written language, is likely to be encoded in people's wordlikeness ratings, so a final analysis might help establish the elements that people are sensitive to when they process unfamiliar letter strings.

Summary

Experiment 8 showed that wordlike pseudohomophone primes facilitated responses to words over unwordlike primes. Although the effect was harder to detect in targets with strongly entrenched spelling patterns (i.e. high frequency words and words with dominant rimes), the experiment gave strong evidence for phonological priming by wordlike pseudohomophones.

5.4 Wordlikeness revisited

What are the orthographic and phonological factors that influence wordlikeness ratings, and, by implication, the processing of unfamiliar letter strings? We have seen that the rime component was an important predictor of wordlikeness ratings, but this is probably not the sole determinant, because items ending in the same orthographic rime elicited different ratings; for example, pseudohomophones *kord*, -0.11, *sord*, 0.43. This is also true of nonwords, so it cannot simply be an effect of feedback from the base word: (*yoind*, -1.25; *foind*, -0.45; *beal*, 1.13; *geal*, 0.38). What other factors might be implicated? Onset plus vowel is one possibility (e.g. Taraban & McClelland, 1987); another is the statistical probability associated with each individual letter and its positional frequency (e.g. Grainger & Jacobs, 1996; Pelli et al., 2003). N must also be considered, although as has been previously stated, N may be a spurious predictor because its influence is confined to low frequency items only, and it also correlates with other variables, particularly the rime (e.g. Andrews, 1989; Peereman & Content, 1995).

Clearly there are numerous variables that could be incorporated into a decomposition of wordlikeness, but since the main consideration is the extent to which orthographic and phonological factors independently contribute to WL judgements, a simple, broad-brush, solution was adopted, by averaging the $\log_{10}$ counts of occurrences in monosyllabic words of the CVC constituents of the pseudohomophones and nonwords. This therefore represents a count of the relative frequencies of a letter-string's orthographic and phonological constituents in terms of onset consonant/s, vowel/s, and final consonant/s. This achieved measures ('Average orthography' and 'Average phonology' in Table 33), which broadly capture the orthotactic and phonotactic probabilistic patterns in English monosyllables. These measures were correlated with the wordlikeness ratings of the new items and those used by previous researchers (both pseudohomophones and nonwords). Since, as we have seen (Section 3.4.1), raters gave lower ratings to items with non-existent rimes than to those with existing rimes, the analysis was carried out separately for items with and without existing rimes.

Table 33
Summary of orthographic and phonological correlations with nonword wordlikeness ratings.

	Average orthography	Average phonology
Nonwords (n = 508)		
Non-existent rimes	.47**	.18*
Existing rimes	.34**	.24**
Pseudohomophones (n = 262)		
Non-existent rimes	.40**	.02 ns
Existing rimes	.29**	.39**

Table 33 shows that the orthography measure correlates more strongly with the wordlikeness ratings for nonwords and pseudohomophones with nonexistent rimes than with items with existing rimes. The effect is reversed for the phonological measure, although the difference is more apparent for the pseudohomophones. To explore this apparent dissociation further, by subjects analysis was carried out. Each individual rater's item scores were correlated with the orthographic or phonotactic measure for that item, and paired t tests were then used to compare the differences in correlation coefficients according to rime status. Table 34 shows that rime status affected raters' responses differently, in that there was a significant difference between correlation coefficients for the orthographic measure for nonwords and for the phonotactic measure for pseudohomophones. It appears that rime is more salient in terms of the orthographic components of nonwords, but more important phonologically for pseudohomophones.

Table 34
t-tests comparing the difference between raters' correlations with orthographic and phonological measures for items with and without existing rimes.

	Average Orthography	Average Phonology
Nonwords		
Mean difference	.10	-.01
t (47)	4.89, p <.001	.653, n.s.
PHs		
Mean difference	.04	-.24
t (59)	.28, n.s.	6.89, <.001

* = *p <.05*, ** = *p <.001*

To further explore the differences in correlations between the phonological measure and wordlikeness ratings, between items t tests were carried out. For items with non-existent rimes, raters were more influenced by phonology for nonwords than for pseudohomophones (mean difference, .11, t (106) = 3.19, p =.002); but for items with existing rimes, the phonological influence was higher for pseudohomophones than for nonwords (mean difference, -.12, t (106) = - 5.66, p <.001).

The analysis shows that readers made use of their orthotactic and phonotactic knowledge when they made wordlikeness judgements, but, in cases where the letter string had an improbable spelling, phonological influence was seriously impaired for pseudohomophones. With wordlike letter strings we see that the phonological measure was stronger for pseudohomophones than for nonwords (Table 33), which may reflect a lexical component, but might equally well emerge from more frequently encountered sublexical patterns. Overall, the indications are that phonological activation is not likely to influence responses to these stimuli unless letter strings reflect spelling probabilities.

Finally, we need to note that although the rime is clearly salient as a predictor of wordlikeness judgements, the existent/non-existent distinction is not totally predictive of wordlikeness. Some items with existing rimes elicited low ratings (*plawk*, -.6, and *kord*, -.11), while some with non-existent rimes were given high ratings (*traip* .56, *carst*, .66). This may be because readers implicitly know that the –

awk rime, although it exists (as in *hawk*), is very low frequency, and that words beginning with *ko-* are rare; on the other hand, although the *–aip* rime is non-existent, it is, given knowledge of *-ait/-ate; -aim/-ame/- ade/-aim* pairs, plausible. It would be useful to see if we can go beyond the rime to unpack some of the contributing factors to the wordlikeness ratings.

Principal component analysis

How might we establish the additional sources of lexical and sublexical knowledge that characterise wordlike and unwordlike letter strings? There are about two dozen additional metrics that might be considered as potential influences on wordlikeness ratings (see Appendix J), but which of these, if any, do readers actually use? One statistical approach to establish the major sources of variance in wordlikeness judgements would be to enter all of these variables into multiple regression, but, since they have a high degree of collinearity, results from this technique are likely to be confounded. To counter this, one might select an appropriate group of variables to use in multiple regression, based on those factors that have been shown to be important in previous research, such as N, bigram frequency, and positional letter frequency. However, while this would simplify matters, this approach means collinearity is still a potential confounding factor, and any potential latent variables might not emerge from the analysis.

What is needed is a method that reduces the large number of inter-correlating predictor variables to a smaller number of orthogonal measures, which can then be used to carry out multiple regression; and this can be accomplished using principal component analysis (PCA). Rather than accounting for, or predicting, the variability in the wordlikeness measure, the technique describes the underlying structure in the data matrix by establishing the interdependence between variables. Similar variables are then grouped together to produce a smaller set of orthogonal components that describe the data matrix; at this point that the researcher needs to interpret the components by bringing theoretical principles and knowledge to bear on the likely characteristics of the grouped variables. Thus, a large number of all potential variables can be included in the analysis, effects are not masked because of their collinearity, and there is the possibility that new combinations of variables that are not captured by single existing metrics may be uncovered.

It was shown above that the most and least wordlike sets of pseudohomophones and nonwords elicited different patterns of responses from raters, so PCA was carried out for each set separately, since a different pattern of components was likely to be established for each group. Factor scores were also calculated to explore how individual nonwords weighted on to the identified components. Finally, multiple regression was carried out using the new components as predictor variables, in order to discover which variables were most predictive of the wordlikeness ratings.
In all four cases, different components emerged from the analysis, which suggests that different sources of knowledge are invoked when raters deal with these groups of stimuli.

Pseudohomophones

With 28 variables entered into the analysis, seven independent components were identified for the 90 most wordlike pseudohomophones, and these accounted for 75% of the variance in the dataset (Table 35). In multiple regression, these seven variables accounted for very little of the variance in the wordlikeness ratings (adj. R^2 = .032), but this is because the predominant influence on these ratings is base word frequency (r (90) = -.2, p = .06). Controlling for base word frequency showed that the seven new variables accounted for rather more variance (adj. R^2 = .24).

Significant predictors were 4 and 5 in Table 35 below; the overall phonology measure, and rime letter position frequency. This second variable is not accurately described as rime spelling per se (which is actually component 2, which does not contribute in multiple regression) but indexes vowels and final consonants separately in terms of their positional frequency. This helps explain why *draip* is an acceptable pseudohomophone, even though *–aip* is a non-existent rime spelling; there are many monosyllables with *–ai-* as vowels in these positions, and there are many 5-letter words that end in *–p*.

Table 35
Components for the most wordlike pseudohomophones

Component	*Example of PH with strong positive component loading (factor score)*	*Example of PH with strong negative component loading (factor score)*
1. Vowel in context of surrounding letters	*ment (2.49)*	*dore (-1.85)*
2. Overall rime spelling	*sware (2.01)*	*draip (-1.94)*
3. Letter positional frequency	*fale (1.61)*	*stoar (-2.45)*
4. **Overall sublexical phonology**	*tite (2.06)*	*gurth (-1.97)*
5. **Rime by letter positional frequency**	*sence (1.55)*	*churp (-1.73)*
6. Bigram frequency	*chane (2.63)*	*yurn (-1.92)*
7. Initial and final letters	*berd (2.62)*	*keap (-1.82)*

The two predictors uncovered in multiple regression correlated negatively with wordlikeness (sublexical phonology, $r(90) = -.42$, $p < .001$; rime letters, $r(90) = -.27$, $p = .011$). This seems to reflect raters' knowledge of the relative likelihood of phonographemic and orthographic patterns. For example, *fawl, tite,* and *dore* all load heavily on to the sublexical phonology component, but they all have relatively low wordlikeness ratings. This indicates raters' knowledge that the sublexical phonology of these items is unlikely to be spelled using those particular sequences of letters.

For the least wordlike pseudohomophones, six independent variables accounted for 62% of the variance in the dataset (Table 36).

Table 36
Components for the most unwordlike pseudohomophones

Component	*Example of PH with strong positive component loading (factor score)*	*Example of PH with strong negative component loading (factor score)*
1. Onset and vowel	*silc (3.69)*	*nyse (-1.26)*
2. Rime phonography	*croo (3.05)*	*woulph (-1.68)*
3. Rime spelling	*phlash (4.51)*	*hownd (-1.01)*
4. Letter positional frequency	*klass(2.77)*	*dryv (-2.18)*
5. Bigram frequency	*broo (3.88)*	*mohr (-1.78)*
6. Unusual letters	*wurk (2.59)*	*fleze (-1.21)*

Multiple regression showed that these variables together predicted approximately 20% of the variability in wordlikeness ratings (adj. $R^2 = .21$). Two separable influences of rime were identified: component 2 refers to the relationship between rime spelling and sound, while component 3 refers to rime spelling alone. However, these were not significant predictors of wordlikeness in multiple regression, which showed that the only predictor of ratings for these unwordlike pseudohomophones was component 4, letter positional frequency. This component showed a strong positive correlation with wordlikeness ratings ($r(90) = .43$, $p < .001$) indicating that higher ratings were given to pseudohomophones where no individual letters occupied an unusual position. Thus, we see that among this unwordlike set of pseudohomophones, items such as *hedj, gayj* and *mylz* were the lowest-rated, presumably indexing the fact that *-j* and *–z* are unlikely final letters in short English monosyllables, while *theem, lofe* and *soope* were among the highest rated. Thus, letter-position knowledge indexed wordlikeness ratings for this group of unwordlike pseudohomophones.

Nonwords

For the nonwords, six components each were extracted for the wordlike and unwordlike group – each described around 84% of the variance in the dataset. For the unwordlike nonwords, six components were identified; multiple regression explained around 24% of the variance in the wordlikeness ratings and showed that the most important contributor was orthographic rime spelling. This clearly indexed orthotactic violations in nonwords such as *leext, nowg, bedj* and *faije*. The second component captured the relationship between the initial letter/s and the following vowel, and indexed unusual letter combinations at the beginning of letter strings, such as *yownd, pryf* and *ghyt.*

For the wordlike nonwords, rime spelling, onset and vowel orthography, mean bigram frequency and overall phonological frequency were established as important components in describing this dataset. However, none of these variables contributed in multiple regression. Among all the possible predictors, only mean bigram frequency correlated with wordlikeness ratings ($r(90) = .25$, $p = .016$). It is possible that ratings for this set of very wordlike nonwords were affected in some way by their orthographic similarity to existing real words, and this would account for the failure of the sublexical component measures to explain the variance in wordlikeness ratings. As noted above, the primary source of influence on ratings for wordlike pseudohomophones was the base word; so it seems likely that letter strings that are orthographically very similar to real words might also activate lexical representations. For example, nonwords like *drean, storn,* and *braim* may have generated orthographically and phonological similar base word representations (*dream, storm, brain*), and this would have affected the wordlikeness ratings. Such an interpretation is consistent with observation that nonwords *raim* and *storn* generated very high error rates in lexical decision (Experiment 3). The actual source of influence for wordlike nonwords is likely to be difficult to pin down; while some nonwords are clearly very similar to a single word, it is more difficult to identify specific lexical items that might affect *pern* and *stroat*; it is possible that multiple, partially-activated lexical representations are activated. Finally, it should be noted that in this, as in all the other analyses, N did not contribute to any of the identified components.

Overall, this analysis has shed further light on the specific components that contribute to the wordlikeness ratings. Unwordlike pseudohomophones and nonwords are similar, in that they are both characterised by orthotactic violations. These violations prevent phonological activation, for both types of stimulus; there is little difference therefore between *hedj* and *bedj*, *paije* and *raije* – none of these items is likely to activate lexical and/or sublexical phonology. The position is slightly different for the wordlike items, in that pseudohomophone ratings are predicted by phonological factors, whereas establishing predictive variables for the wordlike nonwords has been more difficult; it seems likely that a source of reader knowledge about wordlike nonwords is not captured in the metrics investigated so far.

5.5 Summary

Experiment 7 showed that strongly-entrenched rime spellings can affect target responses in lexical decision even at very short presentation times. Experiment 8 showed that wordlike primes are more likely to prime targets than unwordlike primes, and that the source of the priming is phonological, not orthographic. The final wordlikeness analysis showed that orthotactic violations compromise phonological activation, for pseudohomophones and nonwords alike. Therefore, we can conclude that unwordlike letter strings, which are typically characterised by such violations, are not effective primes because they do not activate phonology to the same degree as wordlike primes, which are characterized by familiar spelling patterns. These familiar spelling patterns reflect different sources of knowledge, including onset and vowel combinations, and individual letter knowledge; the rime, although important, is not the only element in readers' knowledge of spelling-sound mappings.

Chapter 6

Of sledgehammers, nuts and dancing angels: What does visual wordlikeness tell us about word recognition?

"Fancy a grown man writing hujus hujus hujus as if he were proud of it it is not English and do not make SENSE"
(Geoffrey Willans, How to be Topp)

The experiments reported in the preceding chapters aimed to use visually wordlike stimuli in a series of naming and lexical decision experiments in order to explore the nature of the visual word recognition system. Pseudohomophone processing was of particular interest because it can shed light on competing models in general, and, more specifically, can illuminate perspectives based on views of 'strong' and 'weak' phonology. Throughout the preceding chapters, evidence has been given to show that participants responded to the orthotactic regularities in both nonwords and pseudohomophones. Therefore, the argument has been that stimuli that are representative of English monosyllabic spelling patterns should be used in experiments, and that there are problems inherent in interpreting findings derived from unwordlike stimuli. However, it could be argued that, since researchers have routinely dealt with sources of potential orthographic confound in stimulus construction, this emphasis was misplaced, and data from carefully-controlled stimuli have already provided a great deal of useful evidence about the visual word recognition system. The attempt to establish measures of wordlikeness, and the sequence of experiments using more wordlike stimuli is therefore equivalent to taking a sledgehammer to crack a nut. It might also be argued that the attempt to establish the effects of wordlikeness was trivial, and not of great theoretical interest; of course *flore* looks more like a word than *phloar*, but this does not tell us anything about the visual word recognition system. The enterprise, therefore, is equivalent to the apocryphal mediaeval debate about the number of angels that could dance on the head of a pin.

These potential criticisms are easily countered. First, analyses reported in Chapter 2 showed that many studies have been based on reports from stimulus sets containing examples of unwordlike items. Subsequent experiments and analyses have demonstrated that unwordlike items are less likely to activate lexical phonology than their wordlike alternatives. This must mean that conclusions, particularly those concerning the status of phonological activation, are potentially compromised in studies based on such stimuli. This must also impact on conclusions drawn more generally about the nature of the visual word recognition system. The view expressed in the preceding chapters is that readers have intuitive knowledge about properties of letter strings, and can express this explicitly in wordlike ratings; this implicit knowledge extends to experimental responses, and can be investigated empirically. An approach that explicitly acknowledges readers' sublexical knowledge is novel. Studies attempting to capture the information that readers have extracted from exposure to print are the exception rather than the rule, and, where they have been reported, have investigated lexical knowledge. For example, subjective frequency ratings for words were collected by Balota, Pilotti, and Cortese (2001), Gernsbacher (1984) and Peereman, Content and Bonin (1998). No attempts to capture adult readers' implicit knowledge of spelling regularities have been reported in the literature.

In general, the value of language users' intuitions as research evidence is rarely acknowledged in psycholinguistics, although the approach is widely used in the discipline of linguistics. Gordon and Hendrick (1997, p. 334) suggest that this is because these data are not seen as valuable because they do not shed light on language processing. Indeed, the prevailing view is that the processes underpinning monosyllabic word and nonword processing can be uncovered on the basis of objective metrics, such as number of letters, letter positional frequency, bigram and trigram frequency, N, and so on. Of course, some or all of these measures may be needed in order to create adequately controlled stimulus sets; indeed, the question of achieving valid test items has dogged researchers for many years and has sometimes been explicitly addressed; one such paper was entitled: "Making up materials is a confounded nuisance" (Cutler, 1981). The plethora of variables to be considered in psycholinguistic experiments seems to increase every year, and even the most tightly-controlled stimulus sets may be subject to experimenter bias (Forster, 2000). Given that there is still no consensus around some of the major questions in word recognition, despite decades of research using carefully-controlled stimuli, this work has addressed the question of whether some of the problems could be associated with stimulus construction, and whether it is possible to make use of readers' implicit knowledge concerning what constitutes a plausible letter string.

The research reported here has shown that readers' intuitive knowledge about orthotactic and graphophonemic sublexical patterns can be reliably expressed in wordlikeness ratings and, contrary to the prevailing view as expressed by Gordon and Hendrick, can illuminate our understanding of the processes involved in reading. Wordlikeness seems to index the sublexical orthotactic and phonotactic knowledge readers have developed through their exposure to written words; it affects responses in a variety of experimental tasks. Therefore, rather than wordlikeness being just another variable to add to the list of controls, it seems possible that it could be treated as a meta-metric which could actually decrease the number of control factors needed, since wordlikeness already incorporates readers' knowledge of bigram and trigram frequency, rime, letter positional frequency, and so on. More importantly, however, the wordlike stimuli can tell us about the word recognition system itself. Some items have been shown to be reliably more wordlike than others, and we have been able to uncover metrics to predict this. We have seen that wordlike and unwordlike items elicit different responses in experiments; and that the sublexical components that drive wordlikeness ratings are different for pseudohomophones and nonwords, and for wordlike and unwordlike stimuli. So the exercise is more than a process that has identified yet another addition to the controls we need to implement in designing experimental stimuli – it can tell us about the mechanisms in the word recognition system itself.

The rest of this chapter summarises the main findings of interest reported in the preceding chapters, and then goes on to establish the importance of visual wordlikeness and to explore the reasons for the widespread, unspoken, assumption that lists of stimuli that are heterogeneous in terms of wordlikeness are acceptable experimental stimuli. The discussion then centres on what the results from more wordlike items have been able to tell us about the two questions that originally prompted the investigation: the broad issue of phonological activation, and the more specific question of which theoretical model best accounts for monosyllabic word and nonword reading.

6.1 Summary of the findings

Stimuli in all the experiments were constructed from real words, on the grounds that, if they were constructed from real word onsets and rimes, they were not likely to generate unusual letter sequences. Comparing wordlikeness ratings for these and previous researchers' stimuli showed that the new items were more wordlike than previous sets. Two naming experiments using the new pseudohomophones and nonwords in pure and mixed lists showed that there was no pseudohomophone effect that could be attributed to the internal workings of the spelling-to-sound system. Ignoring an effect that was attributable to strategies resulting from list composition, it was shown that both pseudohomophones and nonwords were read via the same processes, and these processes did not invoke lexical phonology. However, individual differences between fast and slow readers suggested that skilled readers might activate lexical phonology in naming performance; these findings were in the opposite direction to those reported in previous research. While case mixing abolished such individual differences in naming, wordlikeness continued to exert an effect in both pure and mixed lists, such that more wordlike items were named more quickly. In visual and phonological lexical decision, the usual patterns of responses as regards words, nonwords and pseudohomophones were obtained. A wordlikeness effect was again apparent, although it was stronger in the visual task, with the more wordlike stimuli delaying responses. In the phonological task, the wordlikeness effect was weaker, and there was a high error rate. Initial analysis of the wordlikeness ratings established that the rime was an important source of information in that items with non-existent or low frequency rimes were likely to be given low wordlikeness ratings. The next three experiments therefore explored the role of rime, and established that the locus of the rime effect was lexical. The final experiment used the masked priming paradigm to demonstrate that more wordlike pseudohomophones facilitated lexical decision responses to word targets over less wordlike primes.

Three different tasks were used: naming, lexical decision, and primed lexical decision, and while it is useful to have evidence from different tasks, care must be taken in interpreting the overall effects since different mechanisms may be implicated (see Grainger & Ferrand, 1996, pp. 625-6). For example, the lexical decision task requires a word-check, while the naming task requires an additional, articulatory, stage, and might be hampered by lexical checking. Despite this, it was clearly established that wordlikeness affected responses irrespective of the nature of the task, and is therefore clearly implicated in the processes underpinning visual word recognition. Analysis of the ratings showed that wordlikeness is indexed by a number of different orthotactic and graphophonemic patterns, and therefore represents readers' implicit knowledge of spelling-to-sound patterns in English. Wordlike items are, by definition, very similar to real words and using such items has shown that: there is no pseudohomophone effect in naming (*soke* and *sote* are named in the same time); but that pseudohomophones in visual and phonological lexical decision elicit different responses from nonwords by virtue of their lexical status. Under conditions of masked priming, there is a clear effect of phonology for words primed by wordlike pseudohomophones, but this is not apparent for unwordlike primes.

6.2 Why is wordlikeness important?

Making the familiar strange is a key technique in uncovering cognitive processes: classic examples include exploring face perception using upside-down faces, speech perception using lip movements paired with inappropriate speech sounds, and attentional processes using video clips of gorillas in the midst of basketball players. The underlying principle is that by disabling one or more of the automatic, over-learned, processes or components normally employed in reading, the workings of the system become more accessible to investigation; so, for example, case alternation may disrupt sublexical, graphemic procedures. The aim behind constructing nonwords and pseudohomophones is to construct one group of stimuli where orthography and phonology are both unfamiliar, and a comparison group where orthography is unfamiliar but phonology is not. Using nonwords in general disables the "look-up" processes involved in accessing familiar lexical forms, and using pseudohomophones disables the processes involved in accessing familiar orthographic forms while potentially still permitting activation of familiar phonological forms. Comparisons between the two sets of stimuli can therefore provide useful information about the status of lexical phonology. However, valid conclusions can only be reached if, first, nonwords and pseudohomophones are totally matched to each other apart from on the one variable of interest (in this case, lexical phonology) and, second, if the stimuli have been constructed so as to produce items that are unfamiliar, but not improbable.

This second point is important, because analysis of the wordlikeness ratings (Section 5.4) and Experiment 8 both showed that phonological activation is less likely to occur for unwordlike letter strings. Therefore, while it is the case that phonological effects have been reported in experiments using lists containing unwordlike items (e.g. McCann & Besner, 1988), it is possible that these effects are attributable to a strategy that is different from the processes that are normally used to decode letter strings with familiar grapheme combinations. Since we are primarily interested in the default processes, effects that are attributable to a specific strategy employed to overcome the problems of encoding unwordlike stimuli are not necessarily going to provide useful information. The preceding chapters have shown that it is possible to construct items that are unfamiliar but not improbable; these have been shown empirically to be more wordlike than those used in previous research, and it is therefore likely that participants respond to such items in the way that they respond to most words.

Ironically, one reason why unwordlike stimuli may have been used in previous studies stems from the eminently reasonable view that letter strings must not be created at random, because of the plethora of potential confounding effects. Therefore controls have to be put in place to make the comparison items as similar as possible in all respects apart from the one that is the variable of interest. If this means that the resulting list of items includes *thiphe* (Rastle & Brysbaert, 2006), then that is not perceived as problematic, because everything important has been controlled for. The fact that this letter string looks unusual - and does not easily suggest a pronunciation – is presumably considered irrelevant, and in any case, any individual item effects will be cancelled out by a similar item in the comparison set of stimuli. Does it matter then that some items are unwordlike, as long as there is equivalence across lists? The answer to this, on the basis of the material presented in the previous chapters, is 'yes'. We know that list composition generates different strategies from participants (e.g. Experiment 1); if a list is heterogeneous in terms of subjective ratings of wordlikeness, we cannot be sure what effect that may have on the outcome measure.

Although the notion of 'wordlikeness' has cropped up in the literature from time to time, it has never been operationalised in the same way as in the work reported here. For example, it may be conceptualised as reflecting N (e.g. Davis & Lupker, 2006; Carreiras, Perea, & Grainger, 1997). Davis and Lupker, (Experiment 3) explored the notion that wordlike and unwordlike nonword foils in primed lexical decision would generate different responses to target words, but they operationalised wordlikeness in terms of N. Thus, the prevailing approach has been to impose notions of what characterises letter strings without taking into account readers' top-down existing knowledge; and these notions are embedded in the theoretical approach to the reading system itself. The dominant dual-route approach assumes that a set of rules characterise regular words, and these rules can be applied to create nonwords, which will therefore also be 'regular' – although, as we have seen, these may also sometimes be orthographically very bizarre. As has been shown in the analyses in the previous chapters, such nonwords are unlikely to activate phonology. Therefore it is not surprising that experiments using nonwords with strange orthographies typically produce evidence for a view of the reading system, such as the DRC approach, that prioritises orthographic processing over phonological. The connectionist and resonance approaches offer more fluid, dynamic accounts, and one of the fundamental tenets of this school of thought is that readers develop a knowledge of the distributional properties of printed words. It is not surprising therefore that stimuli from this group of researchers tend to be more wordlike. For example, Seidenberg et al. (1996) introduced the technique of changing real word onsets and rimes to create stimuli, and the salience of rime spelling (an important component of wordlikeness ratings) was established by Herdman et al., (1996) and Vanhoy and Van Orden (2001).

6.2.1 Is N a read herring?

Let us take N as an example of the type of orthographic metric that is frequently controlled for, particularly in approaches characterised by a localist view of the word recognition system. N is a useful example to explore, because its effects are the focus of a major line of research (see Andrews, 1997, for a review). However, the work in the previous chapters strongly suggests that N is potentially a spurious predictor, since there were few significant or systematic effects of N (although it should be acknowledged that N was not explicitly controlled for in the studies reported above, and any analyses were carried out *post hoc*). Since N is a purely orthographic variable and, in part, may index rime frequency (Andrews, 1989; Peereman & Content, 1995), it is likely that it also captures some of the characteristics of wordlikeness, which has also been shown to index rime frequency, among its other components. The key difference between the two metrics is that N is a unidimensional count of orthographic neighbours, while wordlikeness is an abstraction of distributional sublexical orthographic and phonographic properties. N effects are usually interpreted in terms of activation of similar items in the lexicon, and are therefore most strongly associated with a symbolist account. In theory, in a distributed network, N effects could also be defined as emerging from a stable network state (Monsell, Doyle & Haggard, 1989; Van Orden & Goldinger, 1994), but in practice such highly-interconnected networks are likely to represent phonological properties as well, and are therefore more likely to show wordlikeness rather than N effects. Indeed, the computer simulations reported in Chapter 2 showed that the DRC model was sensitive to N but not to wordlikeness; the PDP models showed the reverse effect. Unpicking the relationship between wordlikeness and N would help to shed further light on the word recognition system, but for the purposes of this chapter, detailed analyses are not appropriate. However, since it will be argued that a model of the visual word recognition system that privileges orthographic variables such as N is not an adequate explanation of the empirical data, it is worth considering how we might explore their relationship.

How might we try to establish the relative contributions of N and wordlikeness in nonword processing? Two potential lines of research are suggested. First, Coltheart and Coltheart (2000) found that pseudohomophone rejections in lexical decision were less delayed when the stimulus was not a neighbour of the parent word; they argued that latencies were less delayed because the non-neighbour did not activate the base word, which in turn did not inhibit 'no' responses. Experiment 3 showed that 'no' latencies were less delayed when the stimuli were less wordlike, indicating that it was not necessarily a question of how far the stimulus looked like its base word, but more a question of how far the stimulus looked like ***any*** word. The test therefore would be to directly compare responses to pseudohomophones controlled for N and wordlikeness: the N explanation would predict that *brayn* would delay responses, by virtue of its close orthographic relationship to the base word *brain*, while the wordlikeness explanation would predict that *brane* would delay responses because of its greater orthotactic plausibility; indeed, *brayn*'s unwordlikeness would enable quick rejection on orthographic grounds alone.

The second suggestion relates to Davis and Lupker's claim that wordlikeness could be operationalised in terms of N; they showed that nonword foils with higher N slowed responses to targets in primed lexical decision. While their argument was that the N-wordlike items would activate more orthographic representations, the work reported above suggests that visually wordlike items would have a phonological effect. A quick inspection of Davis and Lupker's stimulus lists revealed that the high-N items, not surprisingly, tended to have immediately recognisable rimes (*rone, rame, kell*) whereas the low-N list tended to have strange rimes (e.g. *rese, vilb, gict*). In terms of wordlikeness, therefore, the high-N list was visually more wordlike than the low-N. The prediction that they explored in that experiment, that nonword foil orthography would affect responses is not at issue; but the argument that the underlying mechanism is orthographic rather than phonographic can clearly be questioned on the basis of the material presented in the preceding chapters. A replication controlling stimuli for orthographic N and wordlikeness could resolve this issue.

6.3 Wordlike stimuli give evidence for 'strong' phonology

One of the key aims of this series of experiments was to investigate the status of lexical phonology and to assess the evidence in favour of the 'strong' or 'weak' approaches summarised by Frost (1998). Overall, the evidence given in the experiments reported above suggests that there is a greater role for phonology in visual word recognition than is suggested by a 'weak' approach as exemplified in the DRC model. We know that wordlike letter strings are more likely to activate phonology than unwordlike (e.g as shown in Experiment 8 and in the statistical analyses of wordlikeness, Section 5.4). Therefore, the failure of previous work to find phonological effects could be attributable to the unwordlikeness of the stimuli. Statistical analysis of the wordlikeness ratings showed that readers based their judgements on the relative frequencies of letters and letter combinations in monosyllabic letter strings, and the unwordlike items were primarily characterised by violations of orthography. Since phonological activation occurs when letter strings are plausible, and is compromised when they are implausible, research that uses unwordlike stimuli might find it difficult to find evidence for a primary role for phonology.

One of Frost's (1998) claims in favour of the 'strong' phonology approach was that there is evidence for phonological recoding even in tasks that should not need it, such as visual lexical decision. Rastle and Brysbaert (2006) countered this claim by pointing out that weak theories had never contested the routine nature of phonological encoding; even the earliest theories had recognised its important influence on visual word processing (e.g. Coltheart et al., 1977). Demonstrating that a phonological code was rapidly assembled, they argued, did not necessarily indicate that it served "any *functional* (my italics) purpose in recognition" (p.100). The alternative view is that not only is a phonological code rapidly assembled, it is functional rather than an epiphenomenon of the system: that is, the system requires phonological encoding in order to operate. Reviewing work by a number of authors, Snowling and Hulme (2005) stated that "a consensus has been reached: phonology is central to word recognition" (p. 5). This view is over-optimistic, because it glosses over the problem of what we mean by "central"; the fact is that research is still divided over whether phonology is the default, or whether phonology plays a role but can still sometimes be bypassed. The evidence from the pseudohomophone effect in visual lexical decision can be interpreted as indicating one of two mutually incompatible processes. Either it represents some spurious activation of phonology during a process that is primarily driven by orthographic processing, or it shows that phonological activation is automatic and mandatory, even when it hampers the task in hand.

The experiments reported above enable us to decide between these two alternatives. In Experiment 7, a primed lexical decision task, phonological recoding was made particularly difficult, since correct decisions could only easily be made if phonological information was suppressed. This resulted in a high error rate, to words as well as pseudohomophones, indicating that phonological recoding was needed in order to make successful orthographic judgements to words like *strode, crane,* and *gripe*. Where all the stimuli look more or less like real words, it becomes very difficult to choose between them on the basis of orthography alone; correct decisions can only be arrived at when phonology is also involved. This would lend support to the view that phonology is an automatic and mandatory component of a fully-functioning word recognition system. Experiment 4 provides a corollary of this finding; in this non-primed phonological lexical decision task, there was also a very high error rate. In circumstances where a useful tactic was to suppress information from orthography, it seemed that the phonological status of the base word was sometimes not enough on its own to generate correct 'yes' responses, resulting in a high number of errors to items like *hoam* and *graid*. Using wordlike stimuli in both experiments made the task difficult; if unwordlike stimuli had been used, the nature of the letter strings would have been a useful clue as to the correct answer. This is not merely a list context effect arising from the particular mix of items (words, nonwords, pseudohomophones); it shows that the word recognition system is highly interactive, and that suppressing sources of either orthographic or phonological knowledge impairs processing. Therefore, phonological processing is not secondary to orthographic – it is functional; in fact, effective functioning requires both phonological and orthographic processing. But using unwordlike stimuli will mask the effect of phonology, and might well produce evidence to foster the conclusion that phonology is of secondary importance.

Sometimes, however, experiments using unwordlike items have given evidence for phonological activation, especially in lexical decision, (e.g. McCann and Besner, 1988) but also in the very brief presentation conditions of masked priming experiments (cf. Rastle and Brysbaert, 2006) and we therefore need to address the fact that phonological effects ***can*** arise from unwordlike stimuli. Given that the argument has been that phonological effects do not emerge for unwordlike items, how can this be? If we acknowledge the fact that phonology is activated even when the nonwords are unwordlike, it indicates how powerful the phonological process actually is. In principle, one would predict that in visual lexical decision where the choice is between unwordlike *gnoys* and real word *noise*, the choice can be made purely on orthographic grounds, and there is no need to invoke phonology; so the fact that a phonological effect is reported in some such cases suggests that the system attempts to engage phonological processes. It is possible that confounds between pseudohomophone and nonword lists can explain these differences, but multiple efforts have been made by researchers in previous studies to control for this problem. As suggested in Section 6.2, a possible explanation is that participants use a different, but reasonably effective, problem-solving strategy to deal with improbable graphemic patterns. Indeed, from time to time, this may be the only option open to us when we are faced with new words that are very unusual. These are often names of people or places; there must be some means of parsing them, or we would throw up our hands in despair when faced with *Kuczaj* and *Bwlch*. For these atypical cases, it might well be the case that a low-level, possibly letter-by-letter, heuristic comes into play in order to generate some coherent phonological representation.

Finally, we need to consider why it is that orthotactic violations result in different patterns of processing. The first possibility is, as suggested above, that it arises as a result of a strategy that involves disabling the phonological mechanism (although this may not always be successful). In lexical decision, it may be possible to decide between real words and unwordlike nonwords on the basis of orthography, and there is no need for phonological activation. Another possibility places the effect within the word recognition system itself rather than as a result of strategic responses, and is therefore of more theoretical interest. This view would hold that the system has, through experience, learned familiar orthography-to-phonology mappings. Such a system can process words and unfamiliar wordlike nonwords efficiently; but unwordlike nonwords, characterised by improbable orthographic patterns, present a processing problem. In order to generate some kind of a response, the system implements an alternative form of processing, such as letter-by-letter reading. This may result in some phonological activation but it is likely to be weakened.

6.4 Which theoretical model best accounts for the findings reported above?

The second main aim of the work reported in the preceding chapters was to explore the nature of the visual word recognition system in terms of the theoretical and computational models that have been proposed to account for observed reading behaviour. In broad terms the initial comparison was between dual-route and single-route approaches. As the work progressed it became clear that the best kind of model to account for all the findings is some kind of interactive model with feedforward and feedback connections between orthography and phonology; the models that best fit this requirement are the MROM-p model (Jacobs et al., 1998) and resonance accounts (Grossberg & Stone, 1986; Gibbs & Van Orden, 1998; Van Orden et al., 1992). However, since the DRC computational model (Coltheart et al., 2001) has bi-directional connections, and since the general dual-route approach has been so influential over the past fifteen years or more, it is not therefore excluded from consideration, although, as we have seen, the primacy it accords to orthography over phonology is likely to weaken its explanatory power. This section briefly discusses the strengths and weaknesses of the main contenders.

Taken together, the findings reported above offer support for a view of a word recognition system that incorporates the following features. First, it is a highly-interactive system, with lexical and sublexical processes operating flexibly according to the task being undertaken. Second, its flexibility includes an ability to deal with unfamiliar input in an efficient manner. This flexibility extends to dealing with improbable input as well, although the mechanisms used for this are probably not the default mechanisms used for more wordlike input. Third, the system makes use of multiple sources of orthographic and phonological information in order to work effectively, but when the task results in suppressing one of these sources of information the system becomes error-prone. Finally, as established in the previous section, phonology is not an optional add-on, but is a functional, mandatory, part of the system.

6.4.1 The DRC model

The DRC model is built according to the 'weak' phonological assumption, so that the GPC route operates more slowly than the lexical route. When tested with the pseudohomophones and nonwords used by McCann and Besner (1987) and Taft and Russell (1992), it produced a good fit to the data (Coltheart et al., 2001). However, when tested with the more wordlike stimuli in Experiment 1, it did not mimic the human data, and read more wordlike items more slowly than unwordlike. Since the model is built so as to process nonwords according to a serial, letter-by-letter process, it has no difficulty with unwordlike items. As suggested above, human readers are also able to deal with unwordlike letter strings, but this is not the default reading process, although it may well be analogous to the slow and rather limited level as currently implemented in the DRC. It is possible that changing the rules embedded in the GPC route might produce greater sensitivity to wordlikeness; at the moment, there is no mechanism that encompasses rimes, which, as we have seen, are an important source of information in processing monosyllabic letter strings. The crucial difference in stimuli *soke* and *hedj* is rime status; human readers process such items differently, but the DRC does not (142, 143 cycles to naming). However, wordlikeness is based on more than just rimes, and the notion of 'rules' is itself not a useful way to conceptualise the multiple grain-sizes of information that appear to be activated when readers respond to letter strings.

In its default setting the DRC computational model is set to read pseudohomophones faster than nonwords, but it can produce the null pseudohomophone effect seen in Experiment 1, by speeding up the GPC route. This effectively turns it into a 'strong' phonological model, but it has the knock-on effect of producing a failure to read exception words correctly (cf. Rastle & Brysbaert, 2006, Simulation 5). Whether this would happen with human readers is a topic for further investigation; do human readers regularise exception words in the context of wordlike pseudohomophones? Lists made up of exception words mixed with pseudohomophones possessing transparent feedforward mappings might result in exception words being given regular pronunciations. On the other hand, the lexical phonology of all list items might militate against such regularisation (cf. the effect of lexical phonology seen in pure lists in Experiment 1). A related phenomenon was investigated by Kay and Marcel (1981), in the context of a similar enquiry; they established that pronunciation of nonwords was affected by preceding words (*nouch – couch/touch*) and they argued that this was better interpreted in terms of a single- rather than dual-process model. But, overall, the DRC model in its present formulation does not offer a good match to the empirical findings regarding naming; one of the reasons for this is arguably because the nonlexical route operates according to a limited set of finite rules. As we have seen, wordlikeness cannot be easily accommodated by a single set of rules since different constituents contribute to high and low wordlike letter strings, and to pseudohomophones and nonwords.

Nevertheless, regarding lexical decision, the explanation given for the pseudohomophone effect is largely the one given by Coltheart et al., in that it emerges by virtue of feedback mechanisms. This is supported to some extent by evidence from Experiments 4 and 7 which suggested that orthographic and phonological information operate bi-directionally; but the important role of phonology is not encompassed within the model. Also, there was no evidence in the experiments reported above for frequency effects arising in the phonological lexicon, as would be expected under the DRC formulation. Rastle and Brysbaert, after a series of simulations with the DRC model, accorded it a slightly more important role for phonology. They suggested that lexical decision is based on the analysis of phonological rather than orthographic representations, but the activation of phonology is constrained by orthographic information. They showed that nonwords and pseudohomophones generated less activation in the phonological lexicon than real words, by virtue of the information encoded in orthography. This echoes remarks that have already been made to the effect that unwordlike letter strings constrain phonological activation, so there is a basis for consensus here. However, given that Rastle and Brysbaert's stimuli included numerous unwordlike items, it would be useful to re-run these simulations with a more wordlike set of stimuli. We have already seen that the model takes longer to name such items, presumably because of competition from activation of spurious lexical candidates, and this might extend to lexical decision so that there was little difference between words and wordlike letter strings.

6.4.2 MROM-p

The MROM-p connectionist computational model (Jacobs et al., 1998) is similar to the DRC model in that it has local representations, but, unlike the DRC, it is single-route. Interactive processes operate between orthographic and phonological sublexical and lexical representations, and these processes can in theory capture the interactions we have seen between multiple sources of information (e.g. Experiments 4 and 7). This is possibly a fruitful model for further investigation since it demonstrated rime feedback effects for words but not for nonwords, analogous perhaps to the human participants in Experiment 5, who generated rime effects for words and pseudohomophones but not for nonwords. In theory, wordlikeness effects might emerge from such a model, but in practice this is unlikely to happen since the input coding scheme is, like the DRC, based on grapheme-to-phoneme conversion 'rules'; in its current formulation it is therefore unlikely to demonstrate the variety of effects we have seen for wordlike and unwordlike nonwords and pseudohomophones.

6.4.3 PDP accounts

As we have seen, the DRC model in its current computational formulation cannot account for the wordlikeness data. Seidenberg and Plaut (2006) argue that this kind of failure to generalise is a "telltale sign that the model has failed to capture the general principles that underlie (reading) behaviour" (p. 29). By contrast, a wordlikeness effect would be an inevitable outcome of a model based on PDP principles, since one of the fundamental tenets of this approach is that readers learn by extracting the statistical regularities of spelling-to-sound relationships that are available to them, and they make use of this knowledge when faced with new items. This knowledge is best expressed as a set of regularities, rather than as a set of immutable rules. The previous chapters have shown that readers do have such knowledge and can express it reliably in wordlikeness ratings, which can be deconstructed into a variety of orthographic and phonological factors which interact systematically (Section 5.4). There is plentiful evidence to suggest that children acquire such knowledge as they learn to read (see Castles & Nation, 2006), and some researchers have even established that children are sensitive to wordlikeness in the sense in which it has been used in this work: for example, Pacton, Perruchet, Fayol and Cleeremans (2001) established that children showed implicit knowledge of orthographic regularities in French. Even first-grade children considered items like *tukke* were more wordlike than *tuuke*, in spite of the fact that they had never seen words containing either *uu* or *kk*. If children are sensitive to distributional information embodied as regularities and patterns in the orthography to which they are exposed, then adults might also be able to demonstrate this sensitivity; and as we have seen, such sensitivity captures not just orthographic but also phonological patterns.

Overall, the data offer support for a dynamic interactive processing system as proposed by the PDP approach and, empirically, one of the connectionist simulations from the Plaut et al. (1996) range of models provided a reasonable approximation to the wordlikeness naming data. The approach taken by the triangle theorists such as Seidenberg, Harm and Plaut has been to produce mini-simulations of one particular aspect of word recognition, rather than to offer an all-encompassing model such as the DRC. Therefore while the data reported above can be interpreted in terms of the theoretical principles underpinning this approach, there is not one single model that can account for all the results taken together. The resonance type accounts (Grossberg & Stone, 1986; Gibbs & Van Orden, 1998; Van Orden et al., 1992) are, broadly speaking, a more fruitful framework in which to interpret the work reported above. As seen by Van Orden and colleagues, the word recognition system is massively interactive, adaptive, and dynamic. In this account, the system self-organises through cooperative feedback between families of nodes (e.g. letter, phoneme, semantic nodes) with special emphasis on letter and phoneme connections. Feedback from phonology rapidly organises perception, with words cohering before nonwords. Gibbs and Van Orden discuss a continuum of nonwords varying in 'legality' (e.g. from *ldfa* to *dilt* to *durt*), which they also term 'wordlikeness'; in this view, unpronounceable letter strings are least wordlike, pseudohomophones are most wordlike. They explain how the system deals with the increasing 'wordlikeness' of such stimuli on the basis of mismatch between bottom-up activation of letter nodes and top-down patterns of activation fed back from phoneme and semantic nodes. "Illegal and legal nonwords entail more mismatch than pseudohomophones because they generate less coherent semantic activity. In turn, illegal nonwords entail more mismatch then legal nonwords because they generate incoherent phonology" (p. 1179). Extrapolating from this, we can argue that such a system will respond to wordlikeness in the way that has been conceptualised and operationalised in the experiments reported above, in a similar way to human readers. A pseudohomophone such as *phloar* would generate a greater degree of mismatch with its base word than *flore*, so it would take longer to name (Experiments 1 and 2), but less time to reject in lexical decision (Experiment 3). *Phloar* would also establish a "generally incoherent phonology" and would therefore be a less successful prime than *flore* (Experiment 8).

Summary

The models that best account for the data reported above are two similar models that invoke the notion of a highly-interactive system of orthographic and phonological lexical and sublexical units, as proposed by Jacobs, Grainger, Ziegler, Van Orden, and colleagues. Because the resonance approach also speaks directly to the way in which a sensitivity to wordlikeness emerges out of exposure to print, it offers the better qualitative match to the results of the experiments reported above.

6.5 Reading disorders

The studies reported in the previous chapters have given evidence for a massively interactive system incorporating orthographic and phonological sources of knowledge, and a functional role for phonology. Although developmental and acquired dyslexia have not been the focus of this work, these findings do have implications for these two reading disorders, and are briefly discussed here.

The word identification problems in children with dyslexia are thought to arise from underlying deficiencies in phonological skills, such as phonological awareness, alphabetic mapping and phonological decoding, and these lead to difficulties in establishing associative bonds between a word's spoken and printed counterparts (Vellutino & Fletcher, 2006). In these children, the "most common and prominent weakness manifests as a deficit in pseudoword reading" (Jackson & Coltheart, 2001, p. 130). This characteristic deficit in nonword reading together with a characteristic phonological impairment suggests either that these children do not develop the sensitivity to the distributional properties of print in the way that typically-developing readers do, or, that they do not develop the rules embedded in a grapheme-to-phoneme conversion system. Developing this further is beyond the scope of this work, but two points may be made here. First, if the reading system is conceptualised as a 'box-and-arrow' system consisting of routes with discrete properties or processes, then explanations for developmental dyslexia are likely to be sought in one of the routes or boxes; so, for example, "we would describe these children as having a deficit at some point along the nonlexical route in their reading system" (Jackson & Coltheart, p. 131). The work reported above suggests that it may not be appropriate to view the reading system in this way; and this is an important issue if the two-route formulation is used as a basis for helping children improve their reading. Second, if reading were just a matter of learning grapheme-to-phoneme correspondences, then we would expect measures of reading difficulty to correlate with measures of intelligence, but deficiencies in lexical skills are apparent in individuals with dyslexia across levels of intelligence (see Steubing, Fletcher, LeDoux, Lyon, Shaywitz, & Shaywitz, 2002, for a meta-analysis). Therefore, reading is more than just learning a set of rules, and if learning to read is understood as a process of developing sensitivity to graphophonemic regularities, this has rather different implications for remediation. (Recent research by Ziegler et al., in press, has suggested that the DRC model can account for different developmental dyslexic profiles on a case-by-case basis; but these authors acknowledge that alternative frameworks and models might also be able to produce similar results (p. 23).

By contrast, acquired dyslexia typically occurs after brain damage such as a stroke, and its manifestations have been interpreted as patterns of preservations and impairments of the modules as conceptualised in the dual route formulation (e.g. Jackson & Coltheart, pp.71-91). Of relevance here is the distinction that has been drawn between surface and phonological dyslexia; patients with the former have difficulty with exception words, while patients with the latter have difficulty with nonwords, thus apparently indicating lesions in either the lexical or the nonlexical route. However, it is possible that this apparent double dissociation emerges as a result of damage to different parts of a highly-interactive system. For example, acquired dyslexia might result from damage to points in shared phonological–semantic space, an explanation based on the triangle model (see Crisp & Lambon Ralph, 2006). On this argument, patients with surface dyslexia would be expected to have co-occurring semantic deficits; and indeed surface dyslexia has been shown to be strongly linked to semantic impairment; with only one exception reported in the literature (Cipolotti & Warrington, 1995). Phonological dyslexia is not linked to semantic impairment, so word reading is not affected but damage to phonological mechanisms results in impaired nonword reading. Most cases of phonological dyslexia also present with identifiable deficits in phonological mechanisms that are not reading-specific. In a special issue of *Cognitive Neuropsychology* (1996), six different papers presented data from 17 patients all of whom had an associated phonological deficit. There are many examples of patients with acquired surface and phonological dyslexia in the literature, but many patients do not have a clear pattern of deficits. Jackson and Coltheart identified at least six subtypes of acquired dyslexia, and Friedman, 1995, differentiated between two types of phonological dyslexia, while Crisp and Lambon Ralph assessed a case series of patients and were unable to identify a "sensible dividing line to separate the participants into distinct groups" (p. 348). If cases are rarely alike in all presenting symptoms, then interpretations of behaviour have to be treated with caution. If the reading system is a highly-interconnected system with multiple processes and sources of knowledge, then this variability in presenting symptoms would be predicted if the system was damaged; but if the reading system has two separate routes with separable functions and processes, then this variability is more difficult to account for, particularly for those patients who present with a mixed pattern of impairments.

However, assuming for the moment that there is a clear subgroup of dyslexic patients whose sole reading difficulty is with nonwords, then it can be argued either that these patients have damage to the GPC route, or that there is damage to some aspect of the process involved in generalising from existing sub/lexical graphophonemic knowledge to new stimuli. There are many case studies that have suggested phonological dyslexia arises from impairment at some point in the GPC route (e.g. MS, Newcombe & Marshall, 1985; WB, Funnell, 1983; ML, Lesch & Martin, 1998). However, there is also evidence to support the notion of a more interactive system. For example, Harley and O'Mara (2006) studied the phonological dyslexic JD and suggested that her impairment resulted from damage to the segmentation procedures that enable efficient nonword reading, a suggestion that is in tune with the view that knowledge about all possible phonological mappings is accumulated through reading experience (Shallice & Warrington, 1987; Shallice et al., 1983). One problem with interpreting evidence from these cases is that patients are often tested on different materials (as also noted by, for example, Patterson, 2000), and this makes comparisons difficult. As has been seen in the previous chapters, non-impaired readers respond to different materials differently, and phonology is activated depending on the orthotactic probability of the letter strings. Since patients are often tested on stimuli from the experimental literature, we can ask to what extent patients have been tested on a homogeneous set of wordlike nonwords. Consider Patterson's stimuli used to test patient CJ; these nonwords were "orthographically well-formed" (p.65), but appear to have varied in wordlikeness. The list included: *werk, roond, wune, muve, teace, leace, broe, gloe, duve, soam, smair, jerm, neen, hoaze* and *wite.* As it happens, most of CJ's errors were typical of those shown by experimental participants. For example, there were lexicalisations (e.g. *werk – week*; *wune – wine*), and there were also responses based on incorrect graphophonemic segmentation (e.g. *soam – so-am*). In Experiment 1, 58% of readers read *kight* as /nIt/ and *chace* was occasionally read as *chance.* The vowel bigram *-oa-* was also sometimes problematic for non-impaired readers (*coad - co-ad; doard – do-ard*). It appears that CJ's responses are qualitatively very similar to those of human readers, but are quantitatively worse (his success rate was 21%), an impairment that seems attributable to his generalised phonological deficit.

However, based on the findings reported in the preceding chapters, it can be suggested that if some of the stimuli used to test patients are unwordlike, it will be difficult to draw conclusions about the underlying mechanisms responsible for patient responses. It has been suggested that unimpaired readers have two mechanisms that operate differently for wordlike and nonwordlike items; wordlike stimuli trigger patterns of graphophonemic activation while unwordlike items generate different processes that are slower and less likely to invoke phonology. It is possible that dyslexic patients will have selective impairments to one or the other process, and that the more challenging task of decoding improbable letter strings might be more susceptible to disruption. It is therefore possible that measures of poor nonword reading may be artificially inflated. In support of this view, we might note that the phonological dyslexic JH was reported as having poor performance on lists of nonwords (Tree & Kay, 2006), but when tested with the stimuli used for

Experiment 1, his naming performance was more accurate than many of the student participants (J. Tree, personal communication, November, 2005). JH made 4.6% naming errors to the pseudohomophone and nonword stimuli, while the student readers made on average 5.5% errors of pronunciation (calculated before trimming). As with CJ, JH also demonstrated lexicalisations, (e.g. *borth – broth)* and also showed effects of nonword length, which is, again, a characteristic of unimpaired readers (Weekes, 1997). From this perspective, JH is no more 'dyslexic' than many student participants; his nonword processing is slower but his accuracy is within the normal range. He has been labelled as a phonological dyslexic on the basis of his very poor nonword reading; but on the basis of his responses to wordlike nonwords, this categorisation must be questioned. It could be argued that, rather than having lost the ability to deal with unfamiliar letter strings, JH's impairment is attributable to a specific, metalinguistic, inability to deal with improbable letter strings.

6.6 Methodological issues and suggestions for future research

Finally, this section addresses a number of potential methodological issues and suggests future lines of research additional to those already outlined. First, there are two questions to do with the construction of the stimuli themselves. One question is whether the wordlike items look more like their base words than their nonword controls. This problem has emerged from time to time in the literature and is clearly a possible confound. The aim was to create stimuli that looked like words as far as possible; but if the pseudohomophones looked more like their base words than the controls, then apparent evidence for phonological effects might actually be orthographic. However, by swapping onsets and rimes (as for Experiments 1 – 4), sublexical graphemic components were very similar in both lists. In Experiment 8 a measure of orthographic similarity was calculated and showed that pseudohomophones and nonwords were equivalent. The second issue is that in order to make more wordlike items, the protocol involved swapping existing onsets and rimes of existing words, which meant that feedback inconsistent items had to be used as the starting points. These items elicit slower responses in lexical decision (Stone et al., 1997) and therefore it could be argued that we have been dealing with an atypical set of items. However, as was pointed out by Ziegler et al., 1997, feedback inconsistent items account for 72% of English monosyllables, so it turns out that this is a rather representative set. Further, we have also demonstrated that items with non-existent rimes can be created as long as they reflect statistical properties of English spelling (e.g. *traip, desp*), so it is likely that there is a body of feedforward consistent items that could be created if required. One advantage that emerges from this approach is that a requirement that stimuli should be wordlike actually limits the number of nonwords and pseudohomophones that can be created, which is useful for replication purposes.

The second issue is that all the experiments have focussed on processing stimuli without reference to semantics. While pseudohomophones may be processed by accessing meaning, (for example, semantic activation in the triangle model), the interpretation of pseudohomophone effects in this work has focussed on processes that are independent of base word meanings. Simplifying matters in this way expedites experimentation and interpretation, but it does ignore the main function of reading, which is to extract meaning. Therefore, future work needs to include an exploration of the role of semantics using wordlike stimuli. Becker et al. (2006) showed that words primed associated words and pseudohomophones (*frog/TOAD/TODE*); consistent with the claims of the resonance approach and with the findings reported above, we might predict that pseudohomophones would also act as associative primes (*tode/FROG/PHROG).* Evidence for the activation of semantic representations would offer strong support for the type of highly interactive model suggested in Section 6.3.4. It would be predicted that unwordlike items would not activate semantics to the same degree as wordlike, and this would offer further support for the argument that phonological representations can be established for unwordlike items, but this is as a result of atypical, problem-solving without reference to normal, semantically-related, lexical and sublexical procedures. Further, although we have seen a critical role for phonology in nonword and pseudohomophone processing, it will not be possible to absolutely determine the role of phonology until its role in word reading is established because, once letter strings are meaningful, it may be that phonology plays a lesser role.

6.7 Conclusion

Reading English words presents a particularly tough challenge by contrast with more 'transparent' orthographies because of the multiple ways in which letters and sounds map on to each other. Representing this variability as a set of rules may (or may not) help beginning readers, but it is not a useful way to conceptualise the mature reading system. In the series of experiments reported here, it has been shown that readers are sensitive to the statistical regularities of their native orthography, and that a reading system with multiple connections between spelling, sound, and meaning has greatest explanatory power. The system is highly interactive, uses multiple grain-sizes of orthographic and phonological knowledge, and when access to these interacting sources of knowledge is impaired, performance is weakened. The system is very powerful and flexible, and it can deal with atypical items if required; but the assumption that processing atypical items can tell us about the normal operations of the reading system is potentially seriously flawed. We ***can*** process *mydz, ehj* and *swhin*, but via an emergency, not a default, mechanism. If the processes we use to deal with these unusual graphemic patterns are taken to be the default mechanisms, this does a serious disservice to the sophisticated and flexible mechanisms that characterise skilled reading. It is crucial therefore that we use stimuli that are valid representations of real words, otherwise experimental results are suspect. Given that many researchers are arguing for the establishment of a set of standard effects in order to evaluate competing models (e.g. Jacobs et al., 1998), we need to be sure that these effects are not driven by confounded stimuli. It doesn't matter whether we count 250 or 205 angels; it does matter whether we write *werds* or *wyrdz*.

References

Allen, P.A., Wallace, B., Weber, T.A. (1995). Influence of case type, word frequency, and exposure duration on visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 914-934.

Andrews, S. (1982). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory & Cognition, 26,* 121-135.

Andrews, S. (1989). Frequency and neighborhood effects on lexical access: Activation or search? *Journal of Experimental Psychology: Learning, Memory and Cognition, 15,* 802-814.

Andrews, S. (1992). Frequency and neighborhood effects on lexical access: Lexical similarity or orthographic redundancy? *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 234-254.

Andrews, S. (1996). Lexical retrieval and selection processes: Effects of transposed-letter confusability. *Journal of Memory and Language, 35,* 775-800.

Andrews, S. (1997). The effect of orthographic similarity on lexical retrieval: Resolving neighborhood conflicts. *Psychonomic Bulletin and Review, 4,* 439-461.

Andrews, S., & Scarratt, D.R. (1998). Rule and analogy mechanisms in reading nonwords: Hough dou peapel rede gnew wirds? *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1052-1086.

Andrews, S., Woollams, A., & Bond, R. (2005). Spelling-sound typicality only affects words with digraphs: Further qualifications to the generality of the regularity effect on word naming. *Journal of Memory and Language, 53(4),* 567-593.

Baayen, R. H., Piepenbrock, R., & van Rijn, H. (1993). *The CELEX Lexical Database* (CD-ROM). Linguistic Data Consortium, University of Pennsylvania.

Balota, D.A., & Chumbley, J.I. (1984). Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 340-357.

Balota, D.A., Pilotti, M., & Cortese, M.J. (2001). Subjective frequency estimates for 2,938 monosyllabic words. *Memory and Cognition, 29(4),* 639-647.

Bailey, T.M. & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighbourhoods? *Journal of Memory and Language, 44,* 568-591.

Baron, J. (1973). Phonemic stage not necessary for reading. *Quarterly Journal of Experimental Psychology, 25,* 241-246.

Beauvois, M.F., & Derouesné, J. (1979). Phonological alexia: Three dissociations. *Journal of Neurology, Neurosurgery, and Psychiatry, 42,* 1125 –1124.

Becker, D.V., Goldinger, S.D., & Stone, G.O. (2006). Perception and recognition memory of words and werds: Two-way mirror effects. *Memory and Cognition, 34,* 1495-1511.

Berent, I. (1997). Phonological priming in the lexical decision task: Regularity effects are not necessarily evidence for assembly. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1727-1742.

Berndt, R. S., Haendiges, A. N., Mitchum, C.C., & Wayland, S.C. (1996). An investigation of nonlexical reading impairments. *Cognitive Neuropsychology, 13,* 763-801.

Besner D., & Johnston, J.C. (1989). Reading and the mental lexicon: On the uptake of visual word identification and production: An examination of four classes of models. In M. Coltheart (Ed.), *Attention and performance:12. The psychology of reading* (pp.201-219). Hillsdale, NJ: Erlbaum.

Bodner, G. E., & Masson, M. E. J. (1997). Masked repetition priming of words and nonwords: Evidence for a nonlexical basis for priming. *Journal of Memory and Language, 37,* 268-293.

Borowsky, R., & Besner, D. (1993). Visual word recognition: A multistage activation model. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 813-840.

Borowsky, R., & Masson, M.J. (1999). Frequency effects and lexical access: On the interpretation of null pseudohomophone base word frequency effects. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 270-275.

Borowsky, R., Owen, W. J., & Masson, M. J. (2002). Diagnostics of phonological lexical processing: Pseudohomophone naming advantages, disadvantages, and base-word frequency effects. *Memory and Cognition, 30,* 969-987.

Borowsky, R., Phillips, J., & Owen, W. (2003, June). *Phonetic decoding processes: On the generalization of the dissociation of frequency and lexicality effects in pseudohomophone naming.* Paper presented at the 13th Annual Meeting of the Canadian Society for Brain Behaviour and Cognitive Science, Hamilton, ON.

Bowey, J. A. (1990). Orthographic onsets and rimes as functional units of reading. *Memory and Cognition, 18,* 419-427.

Bowey, J. A. (1993). Orthographic rime priming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology. Vol 46A,* 247-271.

Bowers, J. (2002). Challenging the widespread assumption that connectionism and distributed representations go hand-in-hand. *Cognitive Psychology, 45,* 413-445.

Bowers J. S., Vigliocco, G., & Haan, R. (1998). Orthographic, phonological and articulatory contributions to masked letter and word priming. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1705-1719.

Brown, G.D.A. (1987). Resolving inconsistency: A computational model of word naming. *Journal of Memory and Language, 26,* 1-23.

Brysbaert, M. & Praet, C. (1992). Reading isolated words: No evidence for automatic incorporation of the phonetic code. *Psychological Research/Psychologische Forschung, 54,* 91-102.

Bub, D., Cancelliere, A., & Kertesz, A. (1985). Whole word and analytic translation of spelling-to-sound in a non-semantic reader. In K. E. Patterson, J. C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 15-34). London: Erlbaum.

Campbell, A.J., & Mewhort, D.J.K. (1980). On familiarity effects in visual information processing. *Canadian Journal of Psychology, 34,* 134 – 154.

Caramazza, A., & Miceli, G. (1990). The structure of graphemic representations. *Cognition, 37,* 243-297.

Carreiras, M., Perea, M., & Grainger, J. (1997). Effects of orthographic neighbourhood in visual word recognition: Cross-task comparisons, *Journal of Experimental Psychology: Learning, Memory and Cognition, 23(4),* 857-871.

Castles, A., & Nation, K. (2006). How does orthographic learning happen? In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*, Hove, UK: Psychology Press.

Cattell, J. McK. (1886). The time taken up by cerebral operations. *Mind, 11,* 220-242.

Chambers, S.M. (1979). Letter and order information in lexical access. *Journal of Verbal Learning and Verbal Behaviour, 18,* 225-241.

Christianson, K. Johnson, R.L., & Rayner, K. (2005). Letter transpositions within and across morphemes, *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1327-1339.

Clark, H.H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour, 15,* 335-359.

Coltheart, M. (2006). Acquired dyslexias and the computational modelling of reading. *Cognitive Neuropsychology. 23,* 96-109.

Coltheart, M., & Coltheart , V. (2000) manuscript in preparation, cited in Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review 108(1),* 204-256.

Coltheart, M., Curtis, B., Atkins, P., & Haller, M. (1993). Models of reading aloud: Dual-route and parallel-distributed-processing approaches. *Psychological Review, 100,* 589-608.

Coltheart, M., Davelaar, E., Jonasson, J.T., & Besner, D. (1977). Access to the internal lexicon. In S. Dornic (Ed.), *Attention and performance V,* (pp.535-555). Hillsdale, NJ: Erlbaum.

Coltheart, M., Laxon, V.J., Keating, G.C., & Pool, M.M. (1986). Direct access and phonological encoding processes in children's reading: Effects of word characteristics. *British Journal of Educational Psychology, 56*, 255-270.

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual-route cascaded model of visual word recognition and reading aloud. *Psychological Review 108,* 204-256.

Content, A. & Peereman, R. (1992). Single and multiple process models of print to speech conversion. In J. Alegria, D. Holender, J. Junca de Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition,* (pp. 213-236). Oxford, England: North-Holland.

Cutler, A. (1981). Making materials is a confounded nuisance: or Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10,* 65-70.

Davis, C., Castles A., & Iakovidis, E. (1998). Masked homophone and pseudohomophone priming in adults and children. *Language and Cognitive Processes, 13,* 625-651.

Davis, C.J. (1999). *The self-organising lexical acquisition and recognition (SOLAR) model of visual word recognition.* Unpublished doctoral dissertation.

Davis, C.J. & Lupker, S.J. (2006). Masked inhibitory priming in English: Evidence for lexical inhibition, *Journal of Experimental Psychology: Human Perception and Performance, 32*, 668-687.

DeFrancis, J. (1989). *Visible speech: The diverse oneness of writing systems.* Honolulu, HI: University of Hawaii Press.

Derouesné, J., & Beauvois, M.F. (1979). Phonological processing in reading: Data from alexia. *Journal of Neurosurgery and Psychiatry, 42,* 1125-1132.

Derouesné, J., & Beauvois, M. F. (1985). The "phonemic" stage in the non-lexical reading process: Evidence from a case of phonological alexia. In K.E. Patterson, J.C. Marshall, & M. Coltheart (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 399-458). London: Erlbaum

Devitt, M. (2006). Intuitions in linguistics. *British Journal of the Philosophy of Science, 57,* 481-513.

Doctor, E. A., & Coltheart, M. (1980). Children's use of phonological encoding when reading for meaning. *Memory and Cognition, 8(3),* 195-209.

Drieghe , D., & Brysbaert, M. (2002). Strategic effecs is associative priming with words, homophones, and pseudohomophones. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 28, 951-961.

Ellis, A.W., Lambon Ralph, M.A., Morris, J., and Hunter, A. (2000). Surface dyslexia: Description, treatment, and interpretation in E. Funnell, (Ed.), *Case studies in the neuropsychology of reading.* Hove, UK: Psychology Press.

Ferrand, L & Grainger, J. (1992). Phonology and orthography in visual word recognition: Evidence from masked non-word priming. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 45A,* 353-372.

Forster, K. I. (2000). The potential for experimenter bias effects in word recognition experiments. *Memory and Cognition, 28,* 1109-1115.

Forster, K. I., & Chambers, S.M. (1973). Lexical access and naming time. *Journal of Verbal Learning and Verbal Behavior, 12,* 627-635.

Forster, K. I., Davis, C., Schoknecht, C., &Carter, R. (1987). Masked priming with graphemically related forms: Repetition or partial activation? *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 39A,* 211-251.

Forster, K. I., & Forster, J.C. (2003) DMDX: A Windows display program with millisecond accuracy. *Behaviour research methods, instruments, and computers, 35,* 116-124.

Frankish, C. R. & Barnes, L. (2008). Lexical and sublexical processes in the perception of transposed-letter anagrams. *Quarterly Journal of Expermental Psychology, 61(3),* 381-391.

Frankish, C. R. & Turner, E. (2007). SIHGT and SUNOD: The role of orthography and phonology in the perception of transposed letter anagrams. *Journal of Memory and Language, 56,* 189-211.

Frost, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123,* 71-99.

Frost, R., Ahissar, M., Gotesman, R., & Tayeb, S. (2003). Are phonological effects fragile? The effect of luminance and exposure duration on form priming and phonological priming. *Journal of Memory and Language, 48,* 346-378.

Frost, R., Kugler, T., Deutsch, A., & Forster, K. I. (2005). Orthographic structure versus morphological structure: Principles of lexical organization in a given language. *Journal of Experimental Psychology: Learning, Memory and Cognition, 31,* 1293-1326.

Funnell, E. (1983). Phonological processes in reading: New evidence from acquired dyslexia. *British Journal of Psychology, 74,* 159-180.

Funnell, E., (Ed). (2000). *Case studies in the neuropsychology of reading.* Hove, UK: Psychology Press.

Gerken, L. & Bever, T.G. (1986). Linguistic intuitions are the result of interactions between perceptual processes and linguistic universals. *Cognitive Science, 10,* 457-476.

Gernsbacher, M.A. (1984). Resolving 20 years of inconsistent interactions between familiarity and orthography, concreteness and polysemy. *Journal of Experimental Psychology: General, 113,* 256-281.

Gibbs, P. & Van Orden, G. C. (1998). Pathway selection's utility for control of word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1162-1187.

Gilhooly, K.J. & Logie, R.H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behaviour Research Methods and Instrumentation, 12,* 395-427.

Glushko, R. J., (1979). The organization and activation of orthographic knowledge in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 5,* 674-691.

Gordon, P.C., & Hendrick, R. (1997). Intuitive knowledge of linguistic co-reference. *Cognition, 62,* 325-370.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Grainger, J., Bouttevin, S., Truc, C., Bastien, M., & Ziegler, J. (2003). Word superiority, pseudoword superiority and learning to read: A comparison of dyslexic and normal readers. *Brain & Language, 87,* 432-440.

Grainger, J., Diependaele, K., Spinelli, E., Ferrand, L., & Farioli, F. (2003). Masked repetition and phonological priming within and across modalities. *Journal of Experimental Psychology: Learning, Memory and Cognition, 29,* 1256-1269.

Grainger, J., & Ferrand, L. (1994). Phonology and orthography in visual word recognition:Effects of masked homophone primes. *Journal of Memory and Language, 33,* 218-233.

Grainger, J., & Ferrand, L. (1996). Masked orthographic and phonological priming in visual word recognition and naming: Cross-task comparisons. *Journal of Memory and Language, 35,* 623-647.

Grainger, J., Granier, J. P., Farioli, F., Van Assche, E., & van Heuven, W. (2006). Letter position information and printed word perception: The relative-position priming constraint. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 865-884.

Grainger, J., & Jacobs, A.M. (1996). Orthographic processing in visual word recognition: A multiple read-out model. *Psychological Review, 103,* 518-565.

Grainger, J., Kiyonaga, K., & Holcomb, P.J. (2006). The time course of orthographic and phonological code activation. *Psychological Science, 17,* 1021-1026.

Grainger, J., Muneaux, M., Farioli, F., & Ziegler, J.C. (2005). Effects of phonological and orthographic neighbourhood density interact in visual word recognition. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology, 58A,* 981-998.

Grainger, J., Spinelli, E., & Ferrand, L. (2000). Effects of baseword frequency and orthographic neighbourhood size in pseudohomophone naming. *Journal of Memory and Language, 42,* 88-102.

Grossberg, S., & Stone, G. (1986). Neural dynamics of word recognition and recall: Attentional priming, learning, and resonance. *Psychological Review, 93,* 46-74.

Harm, M. W., & Seidenberg, M. S. (2001). Are there orthographic impairments in phonological dyslexia? *Cognitive Neuropsychology, 18,* 71-92.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review, 106(3),* 491-528.

Harm, M. W., & Seidenberg, M. S. (2004). Computing the meanings of words in reading: Co-operative division of labor between visual and phonological processes. *Psychological Review, 111,* 662-720.

Herdman, C. M., LeFevre, J-A., & Greenham, S. L. (1994). Implicating the lexicon: base-word frequency effects in pseudohomophone naming. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 575-590.

Herdman, C. M., LeFevre, J-A., & Greenham, S. L. (1996). Base-word frequency and pseudohomophone naming. (1996). *Quarterly Journal of Experimental Psychology, 49A,* 1044-1061.

Holmes, V.M., & Ng, E. (1993). Word specific knowledge, word-recognition strategies, and spelling ability. *Journal of Memory and Language, 32,* 230-257.

Holyk, G.G., & Pexman, P.M. (2004). The elusive nature of early phonological priming effects: Are there individual differences? *Brain and Language, 90,* 353-367.

Howard, D., & Best, W. (1996). Developmental phonological dyslexia: Real word reading can be completely normal. *Cognitive Neuropsychology, 13,* 887-934.

Jackson, N.E., & Coltheart, M. (2001). *Routes to reading success and failure: toward an integrated cognitive psychology of atypical reading.* Hove, UK: Psychology Press.

Jacobs, A.M., & Grainger, J. (1992). Testing a semistochastic variant of the interactive activation model in different word recognition experiments. *Journal of Experimental Psychology: Human Perception and Performance, 18,* 1174-1188.

Kay, J., & Marcel, A. (1981). One process, not two, in reading aloud: Lexical analogies do the work of non-lexical rules. *Quarterly Journal of Experimental Psychology, 33A,* 397-413.

Kinoshita, S., & Lupker, S. J. (2002). Effects of filler type in naming: Change in time criterion or attentional control of pathways? *Memory and Cognition, 30,* 1277-1287.

Kinoshita, S., & Lupker, S. J. (2003). Priming and attentional control of lexical and sublexical pathways in naming: A re-evaluation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 405-415.

Laxon, V.J., Coltheart, V., & Keating, C. (1988). Children find friendly words friendly too: Words with many orthographic neighbours are easier to read and spell. *British Journal of Educational Psychology, 58,* 103-119.

Laxon, V., Masterson, J., Pool, M., & Keating, C. (1992). Nonword naming: Further exploration of the pseudohomophone effect in terms of orthographic neighbourhood size, graphemic changes, spelling-sound consistency and reader accuracy. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18,* 730-748.

Laxon, V.J., Smith, B., & Masterson, J. (1995). Children's nonword reading: Pseudohomophones, neighbourhood size, and priming effects. *Reading Research Quarterly, 30,* 126-144.

Lee, H. W., Kambe, G., Pollatsek, A., & Rayner, K. (2005). The lack of pseudohomophone priming effects with short durations in reading and naming. *Experimental Psychology, 52,* 281-288.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 570-590.

Lukatela, G., Eaton, T., Lee, C.H., Carello, C., & Turvey, M. T. (2002). Equal homophonic priming with words and pseudohomophones, *Journal of Experimental Psychology: Human Perception and Performance, 28,* 3-21.

Lukatela, G., Frost, S.J., & Turvey, M.T. (1998). Phonological priming by masked nonword primes in the lexical decision task. *Journal of Memory and Language, 39,* 666-683.

Lukatela, G., Frost, S.J., & Turvey, M.T. (1999). Identity priming in English is compromised by phonological ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 775-790.

Lukatela, G., & Turvey, M.T. (1994a). Visual lexical access is initially phonological: 1. Evidence from associative priming by words, homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 107-128.

Lukatela, G., & Turvey, M. T. (1994b). Visual access is initially phonological: 2. Evidence from phonological priming by homophones, and pseudohomophones. *Journal of Experimental Psychology: General, 123,* 331-353.

Lukatela, G. & Turvey, M. T. (2000). An evaluation of the two-cycles model of phonological assembly. *Journal of Memory and Language, 42,* 183-207.

Marmurek, H.H.C., & Kwantes, P.J. (1994, November). *Orthographic effects in reading pseudohomophones.* Poster presented at the 35th annual meeting of the Psychonomic Society, St Louis.

Marmurek, H.H.C., & Kwantes, P.J. (1996). Reading words and wirds: Phonology and lexical access. *Quarterly Journal of Experimental Psychology, 49A,* 696-714.

Marshall, J. C., & Newcombe, F. (1973). Patterns of paralexia: A psycholinguistic approach. *Journal of Psycholinguistic Research, 2,* 175-199.

Martin, R.C. (1982). The pseudohomophone effect: The role of visual similarity in nonword decisions. *Quarterly Journal of Experimental Psychology, 34A,* 395-410.

Mason, M. (1978). From print to sound in mature readers as a form of reader ability and two forms of orthographic regularity. *Memory and Cognition, 6,* 568-581.

Mayall, K., & Humphreys, G.W. (1996). Case mixing and the task-sensitive disruption of lexical processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 278-294.

Mayall, K., Humphreys, G.W., & Olson, A. (1997). Disruption to word or letter processing? The origins of case-mixing effects. Journal of Experimental Psychology. *Learning, Memory, and Cognition, 23,* 1275-1286.

Mayall, K., Humphreys, G.W., Mechelli, A., Olson, A., & Price, C.J. (2001). The effects of case mixing on word recognition: Evidence from a PET study. *Journal of Cognitive Neuroscience, 13,* 844-853.

McCann, R.S., & Besner D. (1987). Reading Pseudohomophones: Implications for models of pronunciation assembly and locus of word-frequency effects in naming *Journal of Experimental Psychology: Human Perception and Performance, 13,* 14 – 24.

McCann, R.S., Besner, D, & Davelaar, E. (1988). Word recognition and identification: Do word-frequency effects reflect lexical access? *Journal of Experimental Psychology: Human Perception and Performance, 14,* 693-706.

McCarthy, R., & Warrington, E.K. (1986). Phonological reading: Phenomena and paradoxes. *Cortex, 22,* 359-380.

McClelland, J.L., & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88,* 375-407.

Monsell, S., Doyle, M.C., & Haggard, P.N. (1989). Effects of frequency on visual word recognition tasks: where are they? *Journal of Experimental Psychology: General, 118,* 43-71.

Monsell, S., Patterson, K.E., Graham, A., Hughes, C.H., & Milroy, R. (1992). Lexical and sublexical translation of spelling to sound: Strategic anticipation of lexical status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 452-467.

O'Connor, R.E., & Forster, K.I. (1981). Criterion bias and search sequence bias in word recognition. *Memory and Cognition, 9,* 78-92.

Paap, K.R., Chen, L., & Noel, R.W. (1987, November). *Word recognition: Is the sky falling on top-down processing?* Paper presented at the annual meeting of the Psychonomic Society, Seattle, WA.

Paap, K.R., & Herdman, C. (1998). Highly skilled participants and failures to redirect attention: Two plausible reasons for failing to replicate Paap & Noel's effect. *Journal of Experimental Psychology: Learning, Memory and Cognition, 24(4),* 845-861.

Paap, K.R., Newsome, S.L., & Noel, R.W. (1984). Word shape's in poor shape for the race to the lexicon. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 413-428.

Paap, K.R. & Noel, R.W. (1991). Dual route models of print to sound: Still a good horse race. *Psychological Research, 53,* 13-24.

Paap, K.R., McDonald, J.E., Schvaneveldt R.W., & Noel, R.W. (1987). Frequency and pronounceability in visually presented naming and lexical decision tasks. In M. Coltheart (Ed.), *Attention and performance XII: the psychology of reading* (pp.221-442). Hillsdale, NJ:Erlbaum.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning in real word context: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401-426.

Patterson, K.E., Marshall, J.C., & Coltheart, M. (1985). *Surface dyslexia: Cognitive and neuropsychological studies of phonological reading.* Hove, UK: Lawrence Erlbaum Associates Ltd.

Patterson, K.E. & Morton, J. (1985). From orthography to phonology: An attempt at an old interpretation. In K.E. Patterson, J.C. Marshall, & M. Coltheart, (Eds.), *Surface dyslexia: Neuropsychological and cognitive studies of phonological reading* (pp. 335-359). Hove, UK: Lawrence Erlbaum Associates Ltd.

Peereman, R., & Content, A. (1995). Neighborhood size effect in naming: lexical activation or sublexical correspondences? *Journal of Experimental Psychology: Learning, Memory & Cognition, 21,* 409-421.

Peereman, R., & Content, A. (1997). Orthographic and phonological neighbourhoods in naming: Not all neighbours are equally influential in orthographic space. *Journal of Memory and Language, 37,* 382-410.

Peereman, R., Content, A., & Bonin, P. (1998). Is perception a two-way street? The case of feedback consistency in visual word recognition. *Journal of Memory and Language, 39,* 151-174.

Pelli, D.G., Farrell, B., & Moore, D.C. (2003). The remarkable inefficiency of word recognition. *Nature, 423,* 752-755.

Perea, M., Lupker, S.J. (2003a). Transposed-letter confusability effects in masked form priming. In S. Kinoshita, & S.J. Lupker (Eds). *Masked priming: The state of the art. (pp. 97-120).* New York: Psychology Press.

Perea, M., & Lupker, S.J. (2003b). Does jugde activate COURT? Transposed-letter similarity effects in masked associative priming. *Memory & Cognition, 31,* 829-841.

Perea, M., & Lupker, S.J. (2004). Can CANISO activate CASINO? Transposed-letter similarity effects with nonadjacent letter positions. *Journal of Memory and Language, 51,* 231-246.

Peressotti, F., & Grainger, J. (1999). The role of letter identity and letter position in orthographic priming. *Perception and Psychophysics, 61,* 691-706.

Perfetti, C.A., & Bell, L. (1991). Phonemic activation during the first 40 ms of word identification: Evidence from backward masking and priming. *Journal of Memory and Language, 30,* 473-485.

Perfetti, C.A., Bell, L.C., & Delaney, S.M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language, 27,* 59-70.

Pexman, P.M., Lupker, S.J., & Jared, D. (2001). Homophone effects in lexical decision. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 139-156.

Pexman, P.M., Lupker, S.J., Reggin, L.D. (2002). Phonological effects in visual word recognition: Investigating the impact of feedback activation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28,* 572-584.

Plaut, D.C. (1997). Structure and function in the lexical system: Insights from distributed models of word reading and lexcical decision. *Language and Cognitive Processes, 12,* 765-805.

Plaut, D.C. (1999). A connectionist approach to word reading and acquired dyslexia: Extension to sequential processing. *Cognitive Science, 23,* 543-568.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S., & Patterson, K.E. (1996). Understanding normal and impaired word reading in quasi-regular domains. *Psychological Review, 103,* 56-115.

Plaut, D.C., & Shallice, T. (1994). *Connectionist modelling in cognitive neuropsychology: A case study.* Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Price, C.J., Moore, C.J., & Frackowiak, R.S.J. (1996). The effect of varying stimulus rate and duration on brain activity during reading. *Neuroimage, 3*, 40-52.

Pring, L. (1981). Phonological codes and functional spelling units: Reality and implications. *Perception and Psychophysics, 30,* 573-578.

Rastle, K., & Coltheart, M. (1998). Whammies and double whammies: The effect of length on nonword reading. *Psychonomic Bulletin and Review, 5,* 277-282.

Rastle, K., & Coltheart, M. (1999) Serial and strategic effects in reading aloud. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 461-481.

Rastle, K., & Brysbaert, M. (2006). Masked phonological priming effects in English: Are they real? Do they matter? *Cognitive Psychology, 53,* 97-145.

Rayner, K., Sereno, S.D., Lesch, M.F., & Pollatsek, A. (1995). Phonological cues are automatically activated during reading: Evidence from an eye movement priming paradigm. *Psychological Science, 6,* 26-32.

Rosson, M.B. (1985). The interaction of pronunciation rules and lexical representations in reading aloud. *Memory & Cognition, 13,* 90-99.

Rey, A., Ziegler, J.C., & Jacobs A.M. (2000). Graphemes are perceptual reading units. *Cognition, 75,* B1-B12.

Reynolds, M. and Besner, D. (2005). Contextual control over lexical and sublexical routines when reading English aloud. *Psychonomic Bulletin & Review, 12,* 113-118.

Rubenstein, H., Lewis, S.S., & Rubenstein, M.A. (1971). Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning nad Verbal Behaviour, 10*, 645-657.

Rumelhart, D.E., & McClelland, J.L. (1982). An interactive activation model of context effects in letter perception: II. The contextual enhancement effect and some tests and extensions of the model. *Psychological Review, 89,* 60-94.

Scarborough, D.L. Cortese, D., & Scarborough, H.S. (1977). Frequency and repetition effects in lexical memory. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 1-17.

Schilling, H.E.H., Rayner, K., & Chumbley, J.I. (1998). Comparing naming, lexical decision, and eye fixation times: Word frequency effects and individual differences. *Memory and Cognition, 26,* 1270-1281.

Sears, C.R., Hino, Y., & Lupker, S.J. (1995). Neighbourhood size and neighbourhood frequency effects in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 876-900.

Seidenberg, M.S., Petersen, A., MacDonald, M.C., & Plaut, D.C. (1996). Pseudohomophone effects and models of word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 22,* 48-62.

Seidenberg, M., & Plaut, D.C. (2006). Progress in understanding word reading: Data fitting versus theory building, In S. Andrews (Ed.), *From inkmarks to ideas: Current issues in lexical processing*, Hove and New York: Psychology Press.

Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523-568.

Seidenberg, M.S., & McClelland, J.L. (1990). More words but still no lexicon: Reply to Besner et al. (1990). *Psychological Review, 97,* 447-452.

Seidenberg, M.S., Waters, G.S., Barnes, M.A. & Tanenhaus, M.K. (1984). When does irregular spelling or pronunciation influence word recognition? *Journal of Verbal Learning and Verbal Behaviour, 23,* 383-404.

Siegel, L.S., Share, D., & Geva, E. (1995). Evidence for superior orthographic skills in dyslexics. *Psychological Science, 6,* 250-4.

Smith, F. (1969). Familiarity of configuration vs. discriminability of features in the visual identification of words. *Psychonomic Science, 14,* 261-263.

Snowling, M.J. & Hulme, C. (2005). *The science of reading: A handbook.* Malden MA, Oxford UK: Blackwell Publishing Ltd.

Spieler, D.H., & Balota, D.A. (1997). Bringing computational models of word naming down to the item level. *Psychological Science, 8,* 411-416.

Spinelli, F. (1994). *Rôle de la fréquence des mots de base dans la lecture à haute voix des pseudohomophones.* Unpublished master's dissertation, Université René Descartes.

Stanovich, K.E., Siegel, L.S., & Gottardo, A. (1997). Converging evidence for phonological and surface subtypes of reading disability. *Journal of Educational Psychology, 89(1),* 114-127.

Stuebing, K.K., Fletcher, J.M., LeDoux, J.M., Reid, L.G., Shaywitz, S.E. & Shaywitz, B.A. (2002). Validity of IQ-discrepancy classifications of reading disabilities: A meta-analysis. *American Educational Research Journal, 39,* 469-518.

Stone, G.O., Vanhoy, M., & Van Orden, G.C. (1997). Perception is a two-way street: Feedforward and feedback phonology in visual word recognition. *Journal of Memory and Language, 36,* 337-359.

Stone, G.O. & Van Orden, G.C. (1994). Strategic control of processing in word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 20,* 1248-1268.

Taft, M. (2006). A localist-cum-distributed (LCD) framework for lexical processing, in S. Andrews (Ed.), *From inkmarks to ideas.* Hove, UK: Psychology Press.

Taft, M., & Russell, B. (1992). Pseudohomophone naming and the word frequency effect. *Quarterly Journal of Experimental Psychology, 45A,* 51-71.

Taraban, R., & McClelland, J.L. (1987). Conspiracy effects in word recognition. *Journal of Memory and Language, 26,* 608-631.

Tree, J.J., & Kay, J. (2006). Phonological dyslexia and phonological impairment: An exception to the rule? *Neuropsychologia. 44,* 2861-2873.

Treiman, R. (1984). On the status of final consonant clusters in English syllables. *Journal of Verbal Learning & Verbal Behavior, 23,* 343-356.

Treiman, R.. (1994). To what extent do orthographic units in print mirror phonological units in speech? *Journal of Psycholinguistic Research, 23,* 91-110.

Treiman, R.., & Chafetz, J. (1987). Are there onset- and rime-like units in printed words? In M. Coltheart (Ed). *Attention and performance 12: The psychology of reading. (pp. 281-298).* Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Treiman, R., Goswami, U., & Bruck, M. (1990). Not all nonwords are alike: Implications for reading development and theory. *Memory and Cognition, 18,* 559-567.

Treiman, R., Mullennix, J., Bijeljac-Babic, R., Richmond-Welty, E.D. (1995). The special role of rimes in the description, use, and acquisition of English orthography. *Journal of Experimental Psychology: General, 124,* 107-136.

Treiman, R., & Zukowski, A. (1988). Units in reading and spelling. *Journal of Memory and Language, 27,* 466-477.

Tzur, B., & Frost, R. (2007) SOA does not reveal the absolute time course of cognitive processing in fast priming experiments. *Journal of Memory and Language, 56,* 321-335.

Vanhoy, M., & Van Orden, G.C., (2001). Pseudohomophones and word recognition. *Memory and Cognition, 29,* 522-529.

Van Orden, G.C. (1987). A ROWS is a ROSE: Spelling, sound and reading. *Memory & Cognition, 15(3),* 181-198.

Van Orden, G.C. (1991). Phonologic mediation is fundamental to reading. D. Besner & G.W. Humphreys (Eds). *Basic processes in reading: Visual word recognition.* (pp. 77-103). Hillsdale, NJ, England: Lawrence Erlbaum Associates, Inc.

Van Orden, G.C., & Goldinger, S.D. (1994). Interdependence of form and function in cognitive systems explains perception of printed words. *Journal of Experimental Psychology: Human perception and performance, 20,* 1269-1291.

Van Orden, G.C., Jansen op de Haar, M.A., & Bosman, A.M.T. (1997). Complex dynamic systems also predict dissociations, but they do not reduce to autonomous components. *Cognitive Neuropsychology, 14,* 131-165.

Van Orden, G.C., & Kloos, H. (2005). The question of phonology and reading. In M.J. Snowling & C. Hulme (Eds.), *The science of reading: A handbook* (pp. 61-78). Malden MA and Oxford, UK: Blackwell Publishing Ltd.

Van Orden, G.C., Pennington, B.F., & Stone, G.O. (1990). Word identification in reading and the promise of subsymbolic psycholinguistics. *Psychological Review, 97,* 488-522.

Van Orden, G.C., Stone, G.O., Garlington, K.L., Markson, L.R., Pinnt, G.S., Simonfy, C.M, & Brichetto, T. (1992). "Assembled" phonology and reading: A case study in how theoretical perspective shapes empirical investigation. In R. Frost & L. Katz (Eds.), *Orthography, phonology, morphology, and meaning,* (pp. 249-292). Oxford, England: North-Holland.

Vellutino, F.R., & Fletcher, J.M. (2005). Developmental Dyslexia, In M.J. Snowling, & C. Hulme (Eds.), *The Science of reading: A handbook* (pp 363-378). Malden MA, Oxford UK: Blackwell Publishing Ltd.

Verstaen, A., Humphreys, G.W, Olson, A., & d'Ydewalle, G. (1995). Are phonemic effects in backward masking evidence for automatic prelexical phonemic activation in visual word recognition? *Journal of Memory and Language, 34,* 335-356.

Waters, G.S., & Seidenberg, M.S. (1985). Spelling-sound effects in reading: Time course and decision criteria. *Memory & Cognition, 13,* 557-572.

Weber, R.M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly, 5,* 427-451.

Weekes, B. (1997). Differential effects of number of letters on word and nonword naming latency. *Quarterly Journal of Experimental Psychology, 50A*, 439-456.

Welbourne, S.R. & Lambon Ralph, M.A. (2007). Using parallel distributed processing models to simulation phonological dyslexia: The key role of plasticity-related recovery. *Journal of Cognitive Neuroscience, 19(7),* 1125-1139.

Zevin, J.D., & Balota, D.A. (2000). Priming and attentional control of lexical and sublexical pathways during naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26,* 121-135.

Ziegler, J.C., Castel, C., Pech-Georgel, C., George, F., Alario, F-X, & Perry, C. (in press) Developmental dyslexia and the dual route model of reading: Simulating individual differences and subtypes. *Cognition,* (2007) doi: 10.1016/j.cognition.2007.09.004.

Ziegler, J.C., Jacobs, A.M., & Kluppel, D. (2001). Pseudohomophone effects in lexical decision: Still a challenge for current word recognition models. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 547-559.

Ziegler, J.C., Montant, M., & Jacobs, A.M. (1997). The feedback consistency effect in lexical decision and naming. *Journal of Memory and Language, 37,* 533-554.

Ziegler, J.C., & Perry, C. (1998). No more problems in Coltheart's neighborhood: Resolving neighborhood conflics in the lexical decision task. *Cognition, 68,* 53-62.

Ziegler, J.C, Muneaux, M., Grainger, J. (2003). Neighborhood effects in auditory word recognition: Phonological competition and orthographic facilitation. *Journal of Memory and Language, 48,* 779-793.

Zorzi, M., Houghton, G. & Butterworth, B. (1998). Two routes or one in reading aloud? A connectionist dual-process model. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1131-1161.

Appendices

Appendix A

Pseudohomophone and nonword stimuli used in Experiments 1 and 2

Base word	Pseudohomophone (PH)	Wordlikeness (z score) (controlled for N and base word frequency)
beam	beem	0.42
burst	birst	0.10
boat	bote	-0.51
brief	breaf	0.22
brawn	brorn	-2.96
calm	carm	0.84
chain	chane	0.09
churn	chern	0.13
chirp	churp	1.99
claim	clame	0.25
clear	cleer	-0.02
code	coad	-0.02
cone	coan	-1.31
crane	crain	0.97
deem	deam	0.79
dense	dence	0.82
dirt	dert	-1.44
door	dore	-0.30
drawl	draul	1.09
fail	fale	0.05
feel	feal	0.33
firm	ferm	0.29
foam	fome	-0.47
first	furst	-0.10
goal	gole	0.70
grade	graid	0.36
green	grean	-0.33
hurl	hirl	-1.24
home	hoam	0.38
kale	kail	-2.03
lean	leen	0.14
lurk	lirk	-0.59
most	moast	0.78
note	noat	-0.60
purge	perge	1.23
pope	poap	-1.21
pose	poze	-1.52

preach	preech	-0.74
raid	rade	0.59
same	saim	-0.68
seek	seak	1.77
sense	sence	0.92
shirt	shurt	-0.64
soak	soke	-0.27
soap	sope	-0.07
speed	spede	-1.36
spoke	spoak	0.27
steam	steme	0.57
stole	stoal	2.16
store	stoar	0.90
thought	thort	-0.78
terse	turse	0.02

Nonword	Wordlikeness (z score)	Nonword	Wordlikeness (z score)
bence	1.19	heaf	1.29
bope	-0.36	hirst	1.71
bream	1.19	keme	-2.53
broar	0.66	lence	0.66
brole	0.15	loak	-0.26
burse	1.49	moak	-1.29
cade	0.15	noke	-0.99
chail	-1.09	pame	0.47
chede	-1.71	peem	-0.36
choan	-0.46	pern	0.88
clane	1.19	prain	-0.26
cloap	-0.46	raim	-0.46
coze	-0.99	seech	0.15
crale	0.56	sherm	-0.05
curp	0.88	sirk	-2.12
dade	-0.36	sirl	-0.58
daul	0.05	skarm	-1.61
doast	0.66	soat	2.32
dreer	0.56	spean	0.66
feak	0.47	spome	-1.29
feen	-0.36	stoam	0.05
fert	0.05	storn	1.81
foad	-0.05	stort	0.47
fote	-0.46	surt	-0.26
greal	-0.26	terge	-1.2
gurst	-0.16	thole	0.78

Filler PHs and nonwords for Experiments 1 and 2

Pseudohomophone		Nonword	
beek	meer	baim	mork
berd	myle	beal	mort
bort	nale	burve	moud
chace	nerse	caze	murch
chork	nurve	cheel	murn
cort	praze	chode	nace
croke	purch	crawk	nawk
croud	rane	crawn	pight
dorn	rize	doard	proan
fite	sain	feek	roke
flore	scawn	fleer	rofe
fraim	shole	fote	scur
furn	skait	frane	shate
gane	skeam	gair	skode
gode	snale	gaud	skyle
gote	stawk	geef	snait
haud	stoan	hane	sork
hork	stur	hurn	stirm
keal	swurl	keak	stite
kight	tawk	kerse	storn
lerch	teek	koze	swale
loard	theef	leam	tain
mair	tirm	lerd	thale
meak	tode	merch	turl
meel	wate	mize	wole

Appendix B

Pseudohomophone and nonword stimuli for Experiments 3 and 4

From lists as above apart from replacement of 3 pseudohomophones (brorn, kail, turse) with

Base word	Pseudohomophone	Wordlikeness (z score) (controlled for base word frequency)
goad	gode	0.44
hearse	hurse	0.73
spade	spaid	0.57

Word stimuli for Experiment 3 (visual lexical decision)

bathe	dull	leer	slang
beet	faith	list	sleet
bend	fake	lush	sloth
bill	feed	mark	slurp
block	fell	might	snake
bluff	fish	mind	soup
blush	fist	moon	steep
bound	flat	mould	stork
break	fleet	notch	street
brim	float	part	sunk
bring	fool	peat	swept
bush	fresh	plush	swim
chant	fruit	pool	tank
chart	glass	pore	test
chip	grin	punch	thank
cloth	grit	raft	think
crouch	heart	reed	trite
crush	held	ridge	voice
cute	herd	rind	weld
date	hoof	roof	wide
dock	junk	rough	work
dove	knock	rouse	world
down	lapse	save	worn
drag	lard	seep	write
drone	lawn	silt	youth

Word stimuli for Experiment 4 (phonological lexical decision)

bend	down	lard	soup
block	drag	list	steep
blush	faith	mark	stork

break	fake	mind	street
brim	feed	part	sunk
bring	fell	pore	swim
chart	flat	reed	tank
clink	float	ridge	test
cloth	fresh	save	think
crouch	fruit	seep	voice
crush	grape	slang	weld
date	heart	sloth	work
dove	herd	slurp	world

Appendix C

Wordlikeness (WL) ratings for previous pseudohomophone stimuli

Borowsky, Owen & Masson (2002)

PH	WL (z score)	PH	WL (z score)
boan	0.67	layt	-0.38
bote	0.58	mohr	-0.27
braiv	-0.74	mylz	-1.24
breaz	-0.37	mynd	-0.55
cleen	0.38	nyse	-0.55
coalt	0.79	phlash	-0.61
doun	0.03	pryd	-0.54
drore	0.08	seaks	0.97
dryv	-0.81	soke	0.77
feeld	0.07	spawt	-0.16
flore	1.27	stait	0.42
fyne	-0.24	stroal	0.5
gaim	0.67	swoar	0.92
gyde	-0.06	terhn	-0.87
hedj	-1.35	theem	-0.01
helled	0.43	truhmp	-0.92
hoald	0.51	tule	0.31
hoap	0.68	wehn	-0.68
hoest	-0.36	whyt	-1.04
klass	-0.1	woak	0.35
		wyfe	-0.69

Marmurek & Kwantes (1996)

Low frequency		High frequency	
PH	WL (z score)	PH	WL (z score)
ayre	0.55	blynde	-0.64
brooz	-0.53	brawd	0.47
kord	-0.11	corse	1.37
fayn	-0.02	dore	0.79
gayj	-1.27	helth	0.38
gool	0.59	hurd	0.73
hurse	0.73	kort	0.11
klenz	-0.84	lahf	-0.92
kord	-0.11	looz	-0.84
playg	-0.89	phrunt	-1.09
pynte	-0.20	proov	-0.62
relm	0.56	shoor	0.37
shef	0.16	sorse	0.30

sord	0.43	surch	0.30
spunj	-1.22	swette	-0.78
sware	0.94	thoe	0.00
toom	0.70	throo	-0.49
weerd	-0.60	tung	0.18
worp	0.76	tutch	0.26
wosp	-0.28	werss	-0.93
woulph	-1.08	wunse	-0.24
yurn	0.58	wurck	-0.46
		wurth	0.16

Herdman et al. (1996)

Illegal rimes

Low frequency		High frequency	
PH	WL (z score)	PH	WL (z score)
birn	0.46	burth	0.49
booz	-0.64	chayr	-0.65
caik	-0.76	fayst	-0.65
draip	0.57	fownd	-0.86
feest	0.37	groope	0.40
fraie	-0.57	leest	0.27
groov	-0.94	looz	-0.84
gurth	0.80	maik	-0.31
hownd	-0.67	milc	-0.74
hufe	-0.14	plaie	-0.29
layst	-0.60	rufe	0.55
raije	-0.94	tirn	0.45
silc	-0.73	trayd	-0.76
soope	-0.06	waije	-1.15
spayd	-0.84		

Legal rimes

Low frequency		High frequency	
PH	WL (z score)	PH	WL (z score)
bleet	0.84	bair	0.56
doam	0.83	blud	0.19
focks	0.52	bocks	0.58
fole	0.60	cair	0.06
laice	0.89	dait	0.01
mair	0.82	deigh	0.15
neel	0.33	faice	0.32
pait	0.63	fawl	0.74
peech	0.56	heet	0.47
stane	1.20	keap	0.67
sted	0.45	reech	0.47
weap	1.47	weel	0.45

Seidenberg et al. (1996)

PH	WL (z score)	PH	WL (z score)
baik	-0.33	lerk	0.97
blaim	0.72	lofe	-0.01
brume	0.5	luse	0.47
caim	0.01	meak	1.14
chayn	-0.79	mene	-0.01
crood	0.34	ment	1.06
dreem	0.46	paije	-1.14
durt	0.16	playt	-0.71
feal	0.88	pleed	0.92
ferm	0.74	pruve	0.09
fleze	-0.67	pryde	-0.52
fome	0.6	purch	0.82
furst	0.44	refe	-0.1
grean	0.46	shaym	-0.74
groaz	-0.64	staige	0.35
grype	0.17	stail	0.72
haiz	-0.71	stuk	-0.28
hert	0.31	taim	0.46
joak	-0.34	tayp	-1.11
kome	0.28	thefe	-0.31
leef	0.97	waid	0.71
leep	0.93	weet	0.38
		wurk	-0.42

Taft & Russell (1992)

Low Frequency		High Frequency	
PH	WL (z score)	PH	WL (z score)
broo	-0.65	berd	0.61
carst	0.66	blak	-0.15
daim	0.64	brane	0.68
flud	0.03	cawl	0.65
gayne	-0.68	croo	-0.97
gerd	0.07	fite	0.72
gode	0.44	larst	0.25
hoast	1.30	moov	-0.88
knele	-0.09	poast	0.55
lume	1.22	rayne	0.23
proov	-0.62	rume	0.26
quak	0.15	shaip	0.56
raip	-0.20	stawk	0.80
saik	-0.31	tode	0.53
slane	1.48	traid	1.17
slej	-1.17	treet	0.74

snair	1.61	whele	-0.33
spaid	0.57		
tite	0.77		

Post hoc Bonferroni tests comparing the difference in wordlikeness between the new set of pseudohomophones and those of previous researchers

PH set	Mean diff.	Sig.
Borowsky et al. (2002)	.63	**
Marmurek & Kwantes Low Freq	.65	**
Marmurek & Kwantes High Freq	.65	**
Herdman et al. illegal rimes High Freq	0.9	**
Herdman et al. illegal rimes Low Freq	1.0	**
Herdman et al. legal rimes High Freq	.53	n.s.
Herdman et al. legal rimes Low Freq	-.40	n.s.
Seidenberg et al.	.58	**
Taft & Russell, High Freq	.49	n.s.
Taft & Russell, Low Freq	.40	n.s.

Appendix D

Pseudohomophone, nonword and word stimuli for Experiment 5.

Pseudohomophones

Dominant rime		Subdominant rime	
chare	lept	bleek	hoap
cheak	mame	blite	laim
clame	mawl	boult	mait
crain	molt	breem	poak
croke	pade	chawk	preech
croud	rore	clauze	saif
deam	sain	clowd	scraul
delt	seak	creapt	shair
dence	shole	drane	slane
gause	soke	draul	soal
gole	sope	fealt	stoak
halk	speach	fense	swoar
hawl	spight	flite	taim
kight	strate	glaid	teek
lade	tipe	grype	traid
	wafe		voal

Nonwords (created from PHs)

Dominant rime		Subdominant rime	
bept	lence	chaid	kaul
blole	mipe	chense	laif
blore	pame	clane	loak
brate	prame	cloal	meapt
chawl	rawl	crauze	meech
clafe	scrade	croal	mowd
clope	shoke	croar	pite
crole	slare	daul	sawk
dralk	solt	deek	sealt
dright	soud	dite	shait
feach	stight	gaid	spaim
feam	sweak	goak	spoult
flain	tause	groar	strair
glain	telt	heem	tane
	vade		waim

Nonwords (created from words)

Dominant rime			Subdominant rime
bove	jight	bloard	neight
cale	lerm	breil	pache
cark	malve	caight	phrak
chate	mide	caise	pythe
cile	mithe	cait	ruard
clat	plore	clede	rurve
fask	prage	cuide	serk
flard	scash	fasque	sloor
flate	sirk	hirm	smek
fode	stase	hyle	smoad
frone	steck	kype	sterk
gace	streed	larve	stite
gack	swight	lauge	tase
gerve	torse	lauve	teight
hipe	vord	lource	yase
	whace		yoan

Words

Dominant rime			Subdominant rime
bleed	lord	board	load
brash	nerve	cache	masque
brat	park	case	mauve
cask	phrase	chase	moan
clove	pipe	clerk	plait
code	rate	curve	praise
cone	ride	firm	scythe
face	sage	flak	source
halve	slate	floor	starve
heck	smack	freight	straight
horse	smile	gauge	style
kirk	stale	guard	swede
lace	store	guide	trek
light	term	height	type
lithe	tight	jerk	veil
	yard		white

Appendix E

Transposed letter (TL) nonword and word stimuli for Experiment 6.

Nonwords

TL within rime		TL between onset and rime	
cuase	ngiht	colth	siwft
cuoch	nroth	corwd	siwtch
cuorse	paece	cuhrch	soctch
daerth	piant	durnk	sohuld
dnace	pnuch	firend	solpe
dnuce	prmopt	folor	sotck
drwan	prnit	fornt	spalsh
dtich	ptach	galnce	spehre
feirce	ruond	garnd	sperad
flase	scroch	gorund	spirng
flseh	sldege	gorwth	starin
froce	smduge	palnt	steram
fuoght	snese	pharse	steret
fuond	snkae	samll	sterss
geust	spnoge	sapce	stirct
guant	stnech	sapde	stirp
haealth	strach	satnd	sturt
hdege	taech	satte	sutck
hieght	taost	scarwl	terat
hrose	thrist	scirpt	ternd
htach	tmept	scohol	tewlve
jduge	tnogue	sektch	therat
laest	tuaght	selpt	therw
lnegth	tuoch	separ	tihnk
luanch	vgaue	sepnd	tiwce
luonge	vlave	sewpt	tiwst
mgiht	waelth	shirnk	tuhmp
mnoth	wdith	shurb	turck
mrach	wegiht	shurg	turth
muoth	wtich	sikrt	wertch
	yuong	sitng	wirst

Words

barge	depth	noise	snatch
batch	doubt	please	sought
beach	draft	point	south
birth	dream	praise	sparse
block	drive	prince	square
bread	faint	range	start

brief
build
catch
chair
chase
cheap
cheek
chief
child
choose
chrome
claim
clear
close
cloud
coarse
court
crease
fence
field
flight
flood
fought
frame
fringe
glass
globe
grease
great
group
hearth
hitch
house
hunch
knight
ledge
match
saint
sauce
scene
scheme
score
screen
serve
sheet
short
shrewd
shrill
shrine
shrunk
since
sleeve
slight
slouch
smile
still
stitch
style
taste
teeth
thatch
thorn
thrash
thread
thrill
throat
throne
throng
tight
trench
verse
voice
waste
world

Appendix F

Stimuli for Experiment 7

Prime Subdominant PH	Prime Control	TARGET Word	Prime Dominant PH	Prime Control	TARGET Word
bleek	blemk	BLEAK	chare	chaje	CHAIR
blite	blihe	BLIGHT	crain	cratn	CRANE
boor	bocr	BORE	croud	crovd	CROWD
boult	bcult	BOLT	deam	decm	DEEM
breem	brecm	BREAM	delt	dedt	DEALT
broun	brovn	BROWN	dence	denme	DENSE
caud	cavd	CORD	dore	dorw	DOOR
chawk	chamk	CHALK	fome	fomc	FOAM
creapt	crsapt	CREPT	furn	fsrn	FERN
drane	dranh	DRAIN	gause	gauhe	GAUZE
feer	fecr	FEAR	gode	godn	GOAD
fense	fenre	FENCE	kight	kigdt	KITE
glaid	glajd	GLADE	lade	ladm	LAID
grype	grjpe	GRIPE	lept	lrpt	LEAPT
hoam	hodm	HOME	mame	mamc	MAIM
hoap	hodp	LAME	shert	shnrt	SHIRT
mault	maslt	MALT	shole	sholw	SHOAL
preech	prevch	PREACH	soke	sokr	SOAK
soal	socl	SOLE	speach	spevch	SPEECH
stede	stehe	STEED	strate	stratc	STRAIT
stoak	stomk	STOKE	valt	vslt	VAULT
stroad	strozd	STRODE	vurse	vnrse	VERSE
swoar	swocr	SWORE	wafe	wafc	WAIF

Prime Subdominant Word	Prime Control	TARGET PH	Prime Dominant Word	Prime Control	TARGET PH
boast	boavt	BOST	churn	chsrn	CHERN
cheek	chemk	CHEAK	clause	clauhe	CLAUZE
claim	clajm	CLAME	cloud	clomd	CLOWD
croak	crcak	CROKE	code	cohe	COAD
fault	fanlt	FALT	dome	domc	DOAM
fraud	fravd	FRORD	drawl	drazl	DRAUL
goal	gocl	GOLE	dream	drecm	DREME
hall	haml	HAWL	felt	fslt	FEALT
hawk	hatk	HALK	flight	flipht	FLITE
loam	locm	LOME	lurch	lnrch	LERCH
moult	monlt	MOLT	mate	matn	MAIT
noun	nomn	NOWN	nurse	nwrse	NERSE
paid	pald	PADE	pence	pevce	PENSE

perch	pnrch	PURCH	pert	psrt	PIRT
roar	rocr	RORE	poke	pokm	POAK
sane	sanm	SAIN	reach	repch	REECH
scheme	schemr	SCHEAM	safe	safr	SAIF
screech	scremch	SCREACH	salt	swlt	SAULT
sense	senre	SENCE	scrawl	scraml	SCRAUL
sneer	snecr	SNEAR	share	sharw	SHAIR
soap	socp	SOPE	slain	slajn	SLANE
spite	spitn	SPIGHT	tame	tamw	TAIM
swede	swehe	SWEED	teak	teok	TEEK
toad	tocd	TODE	trade	tradn	TRAID
type	gpe	TIPE	vole	volr	VOAL

Appendix G
Stimuli for Experiment 8

Group 1		Group 2		Group 3	TARGET
PH	WL	PH	WL	Control	
High WL		*Low WL*			
berd	0.35	burd	0.21	spean	BIRD
breef	0.33	brieph	-1.07	theen	BRIEF
brawd	0.02	brord	-0.51	meap	BROAD
kake	-0.10	caik	-0.60	brore	CAKE
carst	0.08	karst	-0.36	pern	CAST
clame	1.18	claym	-0.80	stroat	CLAIM
cleer	0.13	kleer	-0.10	brete	CLEAR
coad	0.61	kode	-0.08	feach	CODE
kone	0.61	coan	0.23	daul	CONE
corse	1.10	korse	-0.57	droze	COURSE
cort	0.30	kort	-0.89	boak	COURT
crain	1.25	crayn	-0.74	joal	CRANE
croo	-1.12	kroo	-1.24	furt	CREW
daim	-0.30	daym	-1.09	pruse	DAME
dait	0.00	dayt	-0.80	barce	DATE
draip	-0.01	drayp	-1.16	gyne	DRAPE
dreem	0.28	dreme	0.08	mirt	DREAM
fite	0.09	phight	-0.09	kunge	FIRM
ferst	0.10	furst	0.05	pide	FIRST
flore	0.73	phlaw	-0.74	hayk	FLOOR
focks	0.67	phox	-0.70	prete	FOX
graid	0.33	grayd	-0.89	doad	GRADE
heet	0.66	hete	0.02	snocks	HEAT
leef	0.72	lefe	-0.24	krup	LEAF
leen	1.09	lene	0.03	dete	LEAN
lerk	0.62	lirk	0.55	loke	LURK
maik	-0.77	mayk	-1.39	fean	MAKE
meen	0.56	mene	-0.08	clud	MEAN
rade	0.91	rayd	-0.93	pheap	RAID
seak	1.31	seke	-0.62	nait	SEEK
shaim	0.05	shaym	-0.91	blore	SHAME
shaip	-0.14	shayp	-1.24	rorl	SHAPE
shert	0.15	shurt	-0.12	caid	SHIRT
spaid	0.70	spayd	-0.69	boap	SPADE
theef	0.09	thefe	-0.43	pame	THIEF
traid	0.86	trayd	-0.82	clope	TRADE
treet	0.17	trete	-0.41	prane	TREAT
weap	1.98	wepe	-0.22	burge	WEEP
wirk	0.05	wurc	-1.08	crale	WORK
werse	-0.05	werss	-1.17	draym	WORSE
wirth	0.64	wurth	-0.03	yeat	WORTH

Low WL		*High WL*			
bayk	-1.08	baik	-0.56	marse	BAKE
blaym	-0.80	blaim	0.83	murch	BLAME
brayn	-0.68	brane	1.13	stoam	BRAIN
kame	0.13	caim	0.28	stirn	CAME
chayn	-1.05	chane	0.91	doil	CHAIN
kleen	0.19	klean	0.39	borge	CLEAN
klenz	-1.34	klense	-0.19	brame	CLEANSE
dert	0.13	durt	0.41	bome	DIRT
dohm	-1.04	doam	0.29	pyte	DOME
dawe	-0.51	dore	0.61	beal	DOOR
phace	-0.23	faice	-0.21	brole	FACE
phale	0.07	fale	1.06	hine	FAIL
forl	-0.34	fawl	0.53	sipe	FALL
pheest	-0.71	feest	0.62	groal	FEAST
pheel	-0.07	feal	1.36	brare	FEEL
fyne	-0.42	phine	-0.08	seech	FINE
fohm	-1.46	fome	0.68	droon	FOAM
phrunt	-0.99	frunt	-0.36	saip	FRONT
gayne	-0.49	gane	0.76	merse	GAIN
gaym	-1.36	gaige	0.37	hane	GAUGE
grene	0.19	grean	0.63	taige	GUIDE
hird	0.51	hurd	1.03	kurp	HERD
knele	-0.86	neel	0.97	stawt	KNEEL
layt	-0.98	lait	0.01	feek	LATE
lahf	-1.47	larf	0.20	fode	LAUGH
lepe	-0.21	leep	0.98	troom	LEAP
purch	1.10	pirch	1.32	blale	PERCH
playt	-1.21	pleight	0.63	hoak	PLATE
rayne	-0.05	rane	1.04	heef	RAIN
saic	-1.36	saik	-0.76	feam	SAKE
saym	-1.26	saim	-0.30	tance	SAME
spunj	-1.71	spunge	1.28	tirl	SPONGE
tayp	-1.15	taip	-0.31	shurm	TAPE
terhn	-1.46	tirn	0.11	frooze	TURN
whele	-0.45	wheal	1.30	dite	WHEEL
whyt	-1.23	wite	0.64	rork	WHITE

PH = pseudohomophone; WL = wordlikeness (rated by experimental participants)

Nonword fillers

prime	TARGET	prime	TARGET	prime	TARGET	prime	TARGET
barce	BARSE	dete	DEAT	hine	HYNE	pide	PYDE
borge	BAWGE	dite	DIGHT	joal	JOLE	rorl	RAUL
beal	BELE	doad	DODE	clud	KLUD	rork	RAWK
burge	BERGE	doil	DOYL	kunge	KUNJ	saip	SAPE
blale	BLAIL	draym	DRAIM	leet	LETE	seech	SEACH
blore	BLAW	droon	DREWN	loke	LOAK	shurm	SHERM
bome	BOAM	droze	DROSE	marse	MARCE	serm	SIRM

boak	BOKE	feek	FEAK	meap	MEEP	snocks	SNOX
boap	BOPE	pheap	FEAP	mirt	MERT	spean	SPEEN
brame	BRAIM	feach	FEECH	murch	MIRCH	stoam	STOME
brare	BRAIR	feam	FEEM	merse	MURSE	stawt	STORT
brete	BREET	furt	FERT	nait	NATE	stroat	STROTE
brole	BROAL	fode	FOAD	pame	PAIM	stirn	STURN
brore	BROAR	frooze	FRUZE	fean	PHEAN	sipe	SYPE
caid	CADE	gyne	GINE	pyte	PITE	tance	TANSE
clope	CLOAP	groal	GROLE	prane	PRAIN	taige	TAYJ
crale	CRAIL	hayk	HAIK	prete	PREET	theen	THEAN
krup	CRUP	hane	HAIN	pruse	PROOSE	troom	TRUME
kurp	CURP	heef	HEAF	punse	PUNCE	tirl	TURL
daul	DAWL	hoak	HOKE	pern	PURN	yeat	YEET

Appendix H

Orthographic similarity (OS) calculation (from Van Orden, 1987, adapted from Weber, 1970's graphemic similarity (GS) measure).

OS = (GS of target foil and category exemplar)/(GS of category exemplar and itself)

where
$GS = 10([50F + 30V + 10C)/A] + 5T + 27B + 18E)$

F = number of pairs of adjacent letters in the same order shared by stimulus pairs
V = number of pairs of adjacent letters in reverse order shared by stimulus pairs
C = number of single letters shared by stimulus pairs
A = average number of letters in stimulus pairs
T = ratio of number of letters in the shorter stimulus to the number in the longer
B = 1 if the first letter of the pairs is the same, otherwise, B = 0
E = 1 if the last letter in the two words is the same, otherwise, E = 0

Appendix J

Potential contributors to wordlikeness ratings

There are various ways in which letter strings can be statistically categorized. The list below lists possibilities for pseudohomophones (nonwords may be characterized in the same way, but excluding potential influences from base words). As well as including predictors such as N and bigram frequency, which have figured largely in previous research, the list also shows alternative ways of statistically representing the relationship between novel and existing letter strings. These measures all correlate with wordlikeness ratings; but they also inter-correlate, presenting a problem for multiple regression analysis.

1. Base word frequency
   - Type or token
   - Raw or log frequency
2. Orthographic neighbourhood (N)
3. Phonological neighbourhood
4. Bigram frequency
   - Minimum
   - Mean
5. Trigram frequency
   - Minimum
   - Mean
6. Letter slot count measures (PHs)

Counts of words with letters sharing same positions as PH
   - First and last letters
   - Average
   - Average excluding first and last letters
   - Minimum frequency (min frequency of letter in any position)
7. Slot counts (base words)
   - Similar counts as above, but for base word
8. Rime measures
   - Ratio of words with same rime spelling as PH compared to words with different rime spelling
   - Ratio of words with the same spelling as the base word compared to words with different rime spelling
   - Ratio of PH spelling to base word spelling, disregarding alternative spellings
9. Onset and vowel measures
   - Ratio of words with alternative onset + vowel spelling as compared to words with same spelling as that of the PH
10. Onset + vowel, plus rime
    - Summed counts of words with orthographic matches to PH spelling
    - Lower of minimum values for rime and onset + vowel calculations
11. Parsed PH measures
    - Initial cluster frequency
    - Final cluster frequency

- Initial + final frequency
- Minimum frequency parsed counts

12. Phonological measures

As the orthographic measures above for rime, onset + vowel, rime plus onset + vowel, but using phonological frequency counts.

ALL MISSING PAGES ARE BLANK

IN

ORIGINAL